Multidimensional Analysis of Public Open Spaces

Urban Morphology, Parametric Modelling and Data Mining

João V. Lopes[1], Alexandra Paio[2], José N. Beirão[3],
Eliana Manuel Pinho[4], Luís Nunes[5]
[1,2] ISTAR, Instituto Universitário de Lisboa (ISCTE) [3] FAUL, Lisbon [4] CEAU-FAUP, Porto [5] Instituto de Telecomunicações, Instituto Universitário de Lisboa (ISCTE)
[1,2,5] {jvsls|alexandra.paio|luis.nunes}@iscte.pt
[3] jnb@fa.ulisboa.pt [4] empinho@gmail.com

Public open spaces (parks, squares and other gathering places) can only be grasped from a simultaneous view of their attributes. In an ongoing Phd research project we propose to overcome the limitations of traditional-descriptive urban morphology methods in dealing with this simultaneity derived from their many shapes, functions, uses and relations within the urban structure. After developing the relations between formal attributes and intangible spatial properties, their identity and proximity may be disclosed by multivariate statistical analysis and data mining techniques. We outline a multidimensional method for the synchronic analysis and classification of the public open spaces departing from a research corpus of 126 Portuguese urban squares, whose analysis is intended to interactively (re)define it. The work done so far is presented, which comprises: (i) firming the concepts, criteria and attributes to extract; (ii) survey on theories, methods and spatial analysis tools and shortcomings identification; (iii) adaptation and/or creation of new methods and tools; (iv) creation of databases from CAD and GIS environments; (v) research on multivariate analysis, data mining and data visualization techniques.

Keywords: *Urban analysis, Urban design, Public open space attributes, Parametric modelling, Data mining*

1. INTRODUCTION

In general the human mind is not able to apprehend more than four dimensions or simultaneous related variables (Halford et al. 2005). Moreover, some of the public space variables are of a configurational and non-discursive nature (Hillier 1996). The combination of several theories and spatial analysis methods with data mining allows overcoming this inability, creating new bottom-up knowledge from the very structure of the data itself. Big data and the increasing of computational power are pushing the limits of well-established urban morphology practices to more complex challenges. The need for greater integration between the various approaches and schools of urban morphology, especially those based on history and geography and the syntactic or structural ones, and between them and the design and planning practice is emphasized in the literature. The creation of meaningful typological descriptions is essential in capturing desired qualities trapped in the urban structure, and so is their classification.

Through computational processing of a CAD digital database of 126 Portuguese urban squares, and a semi-automatic workflow, we retrieve a heterogeneous set of formal and spatial (metric, geometric and topological-syntactic), environmental and perceptual-cognitive attributes, building an open-ended and scalable database for subsequent multivariate analysis and classification. The identification of correlations and the definition of a minimum set of attributes capable of characterizing the spaces are central to the investigation, as well as the enrichment of the urban datasets with qualitative/performance data.

The multidimensional analysis of this *corpus* will enable: (i) to establish a generic method of classification of public spaces, including contemporary spaces that defy formal classifications; (ii) to propose an alternative classification of the examples, based on the expressiveness of the data and *natural* clustering, suggesting new lines of research; (iii) to correlate quantitative, qualitative and performance attributes, supporting a stronger evidence-based urban design; (iv) to comparatively analyse the descriptive-typomorphological and the configurational-structural approaches; and (v) to develop a synthesis capable of combining, through multidimensional analysis, the most significant contributions of each to the understanding and improvement of the quality of public open spaces by design.

2. BACKGROUND

Studies of urban form seldom carry out detailed analyses, and classification proposals, of urban environments and elements at various scales and complexity levels. Investigations range from the global scale of the settlements to the more detailed scale of the block and street elements. Urhahn and Bobic (1994) catalogue neighbourhoods through a mix of quantitative and qualitative descriptions, scale analysis, and different classification themes, while more strictly quantitative approaches try to escape the limitation of a univariate interpretation by relating attributes and classifying examples through clustering in multivariate triangular graphs (Berghäuser-Pont and Haupt 2010).

Space Syntax (Hillier and Hanson 1984) approaches public space by focusing on the determination of representations and syntactic quantifiable measures that expose the rules of the social construction and of the perceptive-cognitive apprehension of spaces. Campos (1997) studies the relationship between patterns of use and the configuration of the urban network by investigating the penetration of axial lines on the space of a group of urban squares in London, and Campos and Golka (2005) the effect of visual fields through visual graph analysis (VGA, Turner et al. 2001) and isovists (Benedikt 1979). Cutini (2003) studies a number of Tuscan historic squares (piazzas), focusing on the relationship between centrality, configuration and visual measures and tries to extend the typical univariate and bivariate space syntax analysis by creating a new compound VGA index that depicts the hierarchy of convex spaces in settlements.

The methods of urban analysis and classification are usually restricted to 3 or 4 variables, so as to respond to human cognitive, perceptive and visual capabilities and, thus, limiting the simultaneous expression of other features that give spaces their individuality. Gil et al. (2012) compile the shortcomings of traditional typomorfological approaches: their time-consuming methods, limiting the amount of morphological examples and dimensions, their relative opacity and subjectivity, their dependence on the analyst's abilities, as well as on geographical and cultural contexts, questioning their reproducibility and generalization. The identified deficiencies can be addressed 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.

Among the different existing techniques, two of the most used and well-established are clustering and principal component analysis (PCA). Uses of these numerical taxonomy (clustering) and dimensionality reduction (PCA) techniques can be found in related disciplines such as geography, GIS and geocomputation, as well as in the humanities, referred to as factorial analysis.

Within urban morphological studies, they support analyses at different scales: from Laskari et al. (2008) study on urban identity through quantifiable attributes on block shapes at the district level, to metropolitan areas. Gil et al. (2012) carry out an unsupervised classification of the urban fabric of two neighbourhoods of Lisbon focusing on street and block urban elements, and point out a possible integration of these techniques in urban design. Comparing *feature recognition* and *clustering* in urban modelling and classification, Chazar and Beirão (2013) point to their potential for extension to the analysis of non-formal spatial relationships, leading to a better understanding of the public space morphology.

Although multidimensional analysis of the urban void are rarer, Laskari (2014) analyses the residual void space inside the blocks in a neighbourhood of Athens, through their local properties and by different points of view (formal, spatial, relational). The syntactic analysis of convex spaces is being enriched by the ongoing work of Beirão et al. (2014), which introduces a new urban void analytic method that considers its tridimensionality and new classification and aggregation logics of its elementary units. This work also points out to the creation of new perceptual-based metrics for future multivariate classification.

3. THE SQUARE AND THE CORPUS

The main object of this research is to study the formal urban square that historically has a leading role in the urban development of Western cities and culture. Within the framework of the open spaces system of the city it is configured as a public space of individual character, distinct from other urban voids, concentrating both aspects of functional and representative order. While changing throughout history, it maintains a symbolic and identitary factor, as well as a stable formal structure over time.

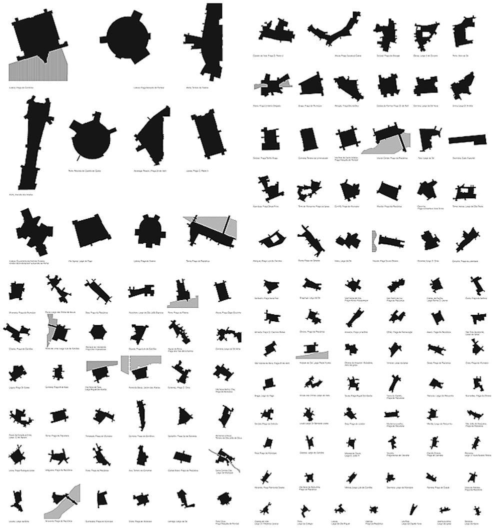

Figure 1
The square in Portugal. The corpus of 126 urban squares. In Coelho et al. (2008).

In order to analyse and classify the public open spaces, it was define a comprehensive *corpus* of 126 Portuguese squares documented in an inventory of public space (Coelho et al. 2008). Its representation is systematic and follows the canons of the Italian school of urban morphology characterized by a strong historical and typomorfological approach (Cataldi et al. 2002). The representation is in digital format and includes, for each of the squares, vectorial 2D drawings of the public space (1/5000 and 1/1000 plans, 1/500 sections, an axonometry) and a simplified 3D CAD model, aerial and street level photos, as well as introductory texts on its location, history and formal and architectural features (Figure 1, above).

4. METHOD

Methodologically, this research has a tripartite basis, collecting concepts and tools from the urban morphology, 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 we use analytic tools from geography (*QGIS*, www.qgis.org), space syntax (*DepthmapX*, UCL) and parametric-algorithmic design (*Rhino/Grasshopper*, McNeel & Assoc.). The advantage of the latter modelling environment in urban analysis at a local scale, over other CAD and GIS platforms, is pointed out by Hanzel (2013). Their associative and interactive 3D nature, rule base approach and data integration flexibility, make it ideal for implementing algorithms and promoting coordination between analysis and design. As the process is explicit, initial criteria and assumptions can be easily changed at a later time. For the multidimensional analysis we resort to *Rapidminer* (RapidMiner GmbH), a popular data mining software based on a visual programming interface (VPI), such as *Grasshopper*.

Data mining, at the intersection of artificial intelligence, machine learning, statistics and database systems, finds patterns and rules in large data sets via an inductive perspective. Its main objectives are prediction (classification of unknown cases and regression) and the discovery of new knowledge (find existing unknown patterns in data). Among the various methods implemented in data mining we intend to explore two of the most established in practice: principal component analysis (PCA) and clustering (k-means algorithm). The PCA analysis determines a smaller set of (artificial) variables that summarize the original data with minimized loss of information and capable of revealing unsuspected relationships. Clustering is an unsupervised learning 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 cases and suggesting interesting hypotheses about relationships (Witten and Frank 2005). These two techniques can be used individually or jointly, with a typical clustering process being preceded by a preliminary dimensionality reduction by PCA.

4.1. Analysis scales and boundaries

In order to adapt the nature of each recorded attribute to the spatial scale and representation which characterize it, five categories of boundaries or spatial aggregation scales are proposed: (1) strict boundary; (2) extended boundary; (3) global boundary, (4) territorial boundary and (5) national boundary.

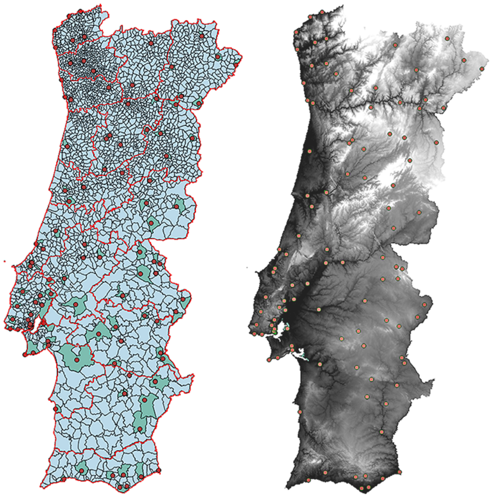

Figure 2
The territorial distribution of the corpus' urban squares in Portugal.

The strict boundary is determined by the projection of the facades plans and the blocks limits that clearly form the space of the square. This determines factors related to shape, perimeter visibility/connectivity and public/private permeability at the more local scale.

The extended boundary is determined by the axes of the streets bordering the adjacent blocks to the previous border. It highlights the aggregation of features related to urban indices, densities and VGA analysis at a local level, while ensuring contextual influence, and corresponds to the district level of aggregation in terms of the *Spacematrix* theory (Berghäuser-Pont and Haupt 2010). The global boundary is determined by the limit of towns and provides information on centrality, intelligibility and synergy features relating local to overall structures. The scale of territorial distribution may be regional (territorial boundary) or national (Figure 2, above). In Table 1 it is shown the relationship between each attribute and its analytical boundary.

4.2. 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 the review of related work and the discussion with specialists. We gather generic morphological, syntactic and environmental descriptions, local and global features from various disciplines that have the urban open space, at its various scales, as the study subject. The heterogeneity of attributes is essential to make the most of the proposed method, for its utility lies in its ability to reveal non obvious correlations and to summarize and visualize big data sets (Witten and Frank 2005).

The attributes were divided into eight groups, or themes, which reflect both the diversity of approaches and scales: (1) Void shape, (2) Vertical plane and permeability, (3) Urban indices and density, (4) Visibility and connectivity (5) Urban system, (6) Use and appropriation, (7) Environment, and (8) Generic labels.

Void shape. Attributes extracted from the two-dimensional representation of the square by its strict boundary (its *figure-ground* diagram), analysed as a polygon. Geometric measures, ratios and properties (shape factors) related to its *elongation*, *rectangularity*, *symmetry*, *spikiness*, and *roughness* are extracted. The curves are approximated by straight segments of minimized deviation.

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 (permeability). The distribution of openings at ground floor is given by a density index per linear meter of façade.

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 based on the indexes defined in the *Spacematrix* theory (Berghäuser-Pont and Haupt 2010) but with a focus on the urban void (square) instead of the built elements.

Visibility and connectivity. Visibility properties, according to three perspectives: (1) VGA limited by the extended boundary, adding only the values of the points inside the square perimeter; (2) the visibility from the exterior of the square, through the calculation of the isovists' overlapping area, created from sets of points outside its strict boundary, and (3) distribution of connectivity/visibility along the square perimeter (Psarra and Grajewski 2001; Laskari et al. 2008).

Urban system. This attribute group focuses on global characteristics of the urban system in which the squares are embedded. We extract the values of the axial lines crossing the square, both global and local, and analyse them in what concerns maximum values of *integration* and *choice*, related to potential *movements to and through* the space (Hillier 1996), average values; and magnitudes and geometric relationships: the sum of the lengths and the averages of the angles formed between pairs of these elements.

Use and appropriation. Attributes which classify buildings adjacent to the square into four classes, according to their use, and register the existence of exceptional buildings, and characteristic elements of the Portuguese urban squares (bandstands and pillories), as well as of urban art, kiosks, fountains and street furniture.

Environment. This group deals with the existence of natural elements (water, green/permeable soil and trees, quantifying their magnitude) and other environmental features such as the percent-

#	Attribute	Code	Boundary	Unit	Data type	Main theory	Authors and References
	VOID SHAPE						
1	**Area**	Vs A	Strict	m2	Real	Shape	
2	**Area/perimeter ratio**	Vs RP	Strict	_	Real	Shape factor	
3	**Nr of perimeter vertices**	Vs NV	Strict	_	Integer	Shape	
4	**Nr of internal islands**	Vs NI	Strict	_	Integer	Shape	
5	**Length of the longest side (angular deviation > 15°)**	Vs LL	Strict	m	Real	Shape	
6	**Aspect ratio (elongation)**	Vs AR	Strict	_	Real	Shape factor	March and Steadman, 1974
7	**Circularity ratio (isoperimetric quotient) (spikiness)**	Vs CI	Strict	_	Real	Shape factor	March and Steadman, 1974
8	**Entropy (symmetry)**	Vs EN	Strict	_	Real	Shape factor/GIS	Claude E. Shannon
9	**Rectangularity ratio (rectangularity)**	Vs RR	Strict	_	Real	Shape factor	Laurini and Thompson (GIS)
10	**Perimeter fractal dimension-Hausdorff dimension (roughness)**	Vs FD	Strict	_	Real	Shape factor	Mandelbrot; Frankauser, 1994
	VERTICAL PLANE, AND PERMEABILITY						
11	**Perimeter/block frontage ratio**	Vp RB	Strict	_	Real	Urban Morphology	
12	**Facade area**	Vp FA	Strict	m2	Real	Urban Morphology	
13	**Maximum height of the facades**	Vp MAH	Strict	m	Real	Urban Morphology	
14	**Mode of the heights of the facades**	Vp MOH	Strict	m	Real	Urban Morphology	
15	**Entropy of the heights of the facades**	Vp EN	Strict	_	Real	Shape factor	Claude E. Shannon
16	**Opening density (opening nr/m frontage)**	Vp OD	Strict	_	Real	Space Syntax	Beirão and Koltsova, 2015
	URBAN RATIOS AND DENSITY						
17	**Nr of adjacent blocks (= nr of adjacent streets)**	De NBL	Strict	_	Integer	Urban Morphology	
18	**Nr of adjacent plots**	De NP	Strict	_	Integer	Urban Morphology	
19	**Nr of adjacent buildings**	De NBU	Strict	_	Integer	Urban Morphology	
20	**Urban square area/adjacent blocks area ratio**	De SBL	Extended	_	Real	Spacematrix	Berghäuser-Pont and Haupt, 2010
21	**Urban square area/adjacent building footprint area ratio**	De SF	Extended	_	Real	Spacematrix	Berghäuser-Pont and Haupt, 2010
22	**Urban square area/adjacent built area ratio**	De SBU	Extended	_	Real	Spacematrix	Berghäuser-Pont and Haupt, 2010
	VISIBILITY AND CONECTIVITY						
23	**Min nr of convex spaces**	Vc MC	Strict	_	Integer	Space Syntax	Hillier and Hanson, 1984
24	**Perimeter mean connectivity (mcv)**	Vc MCV	Strict	_	Real	Space Syntax	Psarra and Grajewski, 2001
25	**Mean differentiation rate (mhv)**	Vc MHV	Strict	_	Real	Space Syntax	Anna Laskari, 2008
26	**Vertical standard deviation of perimeter connectivity (v-value)**	Vc VV	Strict	_	Real	Space Syntax	Psarra and Grajewski, 2001
27	**Horizontal standard deviation of perimeter connectivity (h-value)**	Vc HV	Strict	_	Real	Space Syntax	Psarra and Grajewski, 2001
28	**Area of overlapping isovists from the street (visual exposure)**	Vc VS	Extended	_	Real	Isovist	Benedickt, 1979; Batty, 2001
29	**Urban square area/major isovist (360°) area ratio**	Vc PI	Extended	_	Real	Isovist	Benedick, 1979
30	**Area of the visual integration core (visual integration >90%)**	Vc CO	Extended	_	Real	Space Syntax VGA	Turner, Doxa et al, 2001
31	**Mean visual connectivity (less prone to edge effect)**	Vc COM	Extended		Real	Space Syntax VGA	Turner, Doxa et al, 2001
32	**Mean visual clustering coefficient (less prone to edge effect)**	Vc COE	Extended		Real	Space Syntax VGA	Turner, Doxa et al, 2001
	URBAN SYSTEM						
33	**Mean width of confluent streets**	Us SW	Extended	m	Real	Urban Morph./Design	
34	**Sum of the confluent axial lines length**	Us AL	Global	m	Real	Space Syntax	Hillier, 1999
35	**Mean angle between confluent axial lines**	Us ANL	Global	Degree	Real	Space Syntax	Hillier, 1999
36	**Maximum of the confluent axial lines global integration**	Us GI	Global	_	Real	Space Syntax	Hillier and Hanson, 1984
37	**Maximum of the confluent axial lines local integration (radius3)**	Us GI3			Real	Space Syntax	Hillier and Hanson, 1984
38	**Maximum of the confluent axial lines global choise**	Us GC	Global	_	Real	Space Syntax	Hillier and Hanson, 1984
39	**Maximum of the confluent axial lines local choise (radius3)**	Us GC3			Real	Space Syntax	Hillier and Hanson, 1984
40	**Mean intelligibility (local connectivity/global integration)**	Us IN	Global	_	Real	Space Syntax	Hillier and Hanson, 1984
41	**Mean synergy (local (radius3) integration/global integration)**	Us SY	Global	_	Real	Space Syntax	Hillier and Hanson, 1984
	USE AND APPROPRIATION						
42	**Adjacent building use (Church/historic building; Public building/ Urban equipment; Comerce/Services; Residential)**	Uu BU	Strict	_	Polynom.	Land use	
43	**Urban elements (Bandstand; Pillory; Benches; Kiosk; Urban art; Fountain)**	Uu EL	Strict	_	Polynom.	Urban design/Usage	William H. Whyte; Jan Gehl
	ENVIRONMENT						
44	**Natural elements (Water; Green space; Trees)**	En NE	Strict	_	Polynom.	Site analysis/Landscape	Edward T. White, 1983
45	**Mean sky component (sky hemisphere visiblity %)**	En SK	Strict	%	Real	Site analysis/Solar	Jacques Teller, 2003
46	**Area in constant shadow**	En SH	Strict	_	Real	Site analysis/Solar	Edward T. White, 1983
47	**Global solar orientation (wheghted segment contribution)**	En SO	Strict	_	Real	Site analysis/Solar	Edward T. White, 1983
48	**Maximum topographic slope**	En SL	Strict	%	Real	Site analysis/Landscape	Edward T. White, 1983
	GENERIC LABELS						
49	**Toponymic type (square, churchyard, circus, etc)**	Ge TO	Territory	_	Polynom.	Urban Morphology	
50	**Geografic/climatological region**	Ge GE	Territory	_	Polynom.	Geography	
51	**Latitude (north/south differentiation)**	Ge LA	Territory	Degree	Real	Geography	
52	**Longitude (coast/inland differentiation in Portugal)**	Ge LO	Territory	Degree	Real	Geography	
53	**Site elevation**	Ge EL	Territory	m	Real	Geography	
54	**Date (first historical reference: century)**	Ge DA	Territory	_	Date	Urban Morph./History	
55	**Population density at square civil parishes' level**	Ge PD	Territory	_	Real	Demography	

Table 1
Attribute table. The eight classes of attributes displaying metadata, main theories, authors and references.

age of visible sky area, solar orientation, permanent shadow areas and 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.

Generic labels. These are essentially attributes related to geographical factors and territorial distribution (which may be understood as an expression of socio-cultural difference), or some sort of *a priori* classification (*labelling*), whose correlation with the data can be tested or *learned*. Thus, for example, the physical distance which can be understood as cultural difference, may be correlated with the abstract distance between examples determined by data mining processes (Hanna 2009).

A full list of attributes and their brief description can be found in Table 1 (above). As the *corpus* analysis progresses the definition of the proposed attributes may change and be optimized. In order to increase the expressiveness of the information contained in the data its correlation and type of statistical aggregation shall be attested by EDA processes and by the early modelling of the data mining algorithm.

4.3. Workflow

The proposed workflow includes: (1) preparation of the digital bases and building of CAD models; (2) construction of algorithms for analysis; (3) analysis of the models, attribute extraction and database storage; (4) data mining analysis; (5) representation and mapping; and (6) final results interpretation.

The preparation of the digital bases implies the extraction of the relevant information from the original CAD files in order to provide a basis for the construction of axial, two-dimensional and three-dimensional models. The latter are introduced into the CAD modeller *Rhinoceros*, where they are analysed by algorithmic definitions built in *Grasshopper*; and the former in *DepthMapX* where they are subject to axial and visual syntactic analysis. The extracted data from the parametric modeller is stored in a relational database (*PostgreSQL*) through a data pipeline created by the *Slingshot!* component (Nathan Miller) and *SQL* queries. The *DepthmapX* data is exported in text format, compiled and stored into the same database. This database may be accessed and updated by the GIS program, used in the extraction of geographic attributes at territorial scale, and the data mining analysis program. The database is thus an open-ended, bidirectional and scalable central repertoire, streamlining the workflow. The visualization of statistical information will be made with the *Rapidminer* data visualization tools, and their spatial mapping in *Rhino/Grashopper* and *QGIS*. The final step will be supported by the experts of the morphological study in which this one is based, whose contribution is essential in interpreting the results.

5. ANALYTICAL ALGORITHMS

The values of the local attributes are calculated using the prepared 2D and 3D models of the squares and the parametric/algorithmic modelling environment *Rhino/Grasshopper*. Some of the simple attributes (lengths, areas, counts) are extracted directly without further processing, and the composed attributes (ratios, densities, shape factors) are calculated within the database using the previous ones. We briefly describe some of the developed algorithms requiring more complex and/or specific modelling (Figure 3):

Fractal dimension (Hausdorff dimension by a *coast of Britain-like process*). We use the Hausdorff dimension, one of the various definitions of fractal dimension, to assess the complexity of the square perimeter. Its change in detail with scaling is given by the estimated limit of the slope for the regression line of a *log-log* graph of *size vs scale* (Frankhauser, 1994). In our model, *size* is the number of equal segments (rulers) used to measure the perimeter, which form a polyline whose vertices slide on the perimeter curve, and *scale* is the length of these rulers. The approximation to the Hausdorff dimension, entirely done in *Grasshopper*, results from eight iterations on a scale between 0.5m e 64.0m, doubling the length of the rulers in each iteration.

Perimeter connectivity. Based on work by Psarra and Grajewski (2001) and Laskari et al (2008), we implemented in *Grasshopper* a model that allows the characterization of shapes by analysing the connectivity of their perimeters. The identified measures describe local aspects, and allow the definition of shapes as *patterns of stability and differentiation, rhythm and repetition* and their understanding *beyond the conventional characterization of its geometric order* (Psarra and Grajewski, 2001). The models created in *Grasshopper* have a typical resolution of 1.5 m and calculations are made integrally in the modelling environment.

Shape diameters, radii and diagonals. The outlined algorithm is capable of finding the major segment inscribed within the perimeter of the square, and the major and minor ones perpendicular to it. Deals with convex and concave spaces and is sensitive to the existence of *islands*. The algorithm also calculates the radii centroid-vertices, and variable radial sampling from a selected point. These values allow the determination of attributes related to shape factors whose calculations are mainly taken from the fields of GIS and image analysis.

Internal and external isovists. Isovist fields. Strictly geometrical 2D isovists (without radial sampling) are created. The distinction between isovist polygon segments representing obstacles and occlusion lines is made and measures such as occlusion and drift can be made rigorously. This allows an accurate and interactive control of isovists location and the calculation of properties based on geometric features. Isovist fields can be created step-by-step for points in a grid, using *Grasshopper*'s animated sliders as a simple *foreach-loop* function. However, it is computationally expensive and not presented as a substitute for isovists by radial sampling, or their calculation on specialized programs.

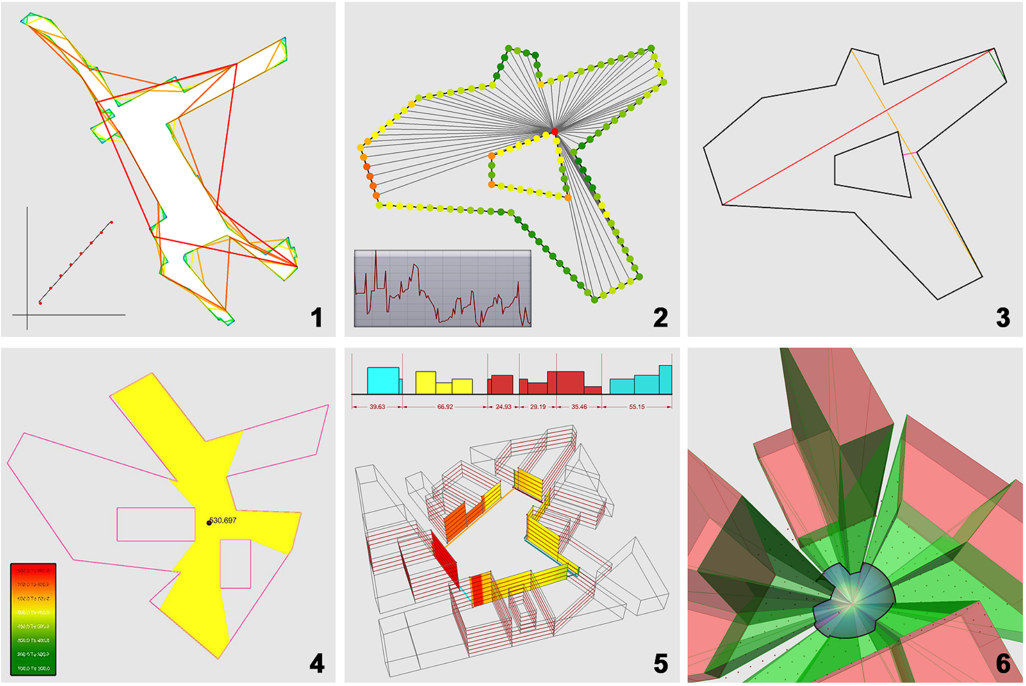

Figure 3
Algorithms in Rhino/Grasshopper. 1.Fractal dimension; 2.Perimeter connectivity; 3.Shape diameters; 4.Parametric isovists; 5.Facades, building heights and solar orientation mapping; 6.Sky factor

Facades, building heights and solar orientation. 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' heights are characterized by the maximum, mode and entropy values of its distribution, respectively accounting for the presence of landmarks, a predominant height and its overall variability. 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 good solar orientation, expressed as a potential since it considers the entire perimeter even if not built.

Sky factor. This algorithm calculates the percentage of the visible area of the sky, from a point, without discretizing the sky vault. The visible surfaces are selected, extruded to that point and subtracted from a hemispherical surface representing the sky. Intercepting the resulting surface with another one representing the solar movement in a given period, and for that latitude, insolation can be approximately calculated. This algorithm records the average percentage of the visible area of the sky from the square and the area permanently in shadow using a square grid of points (sensors) spaced 3.0 m. The calculation is done step-by-step for each of the points as described above in the construction of isovists.

6. CONCLUSION AND FUTURE WORK

The advantages of a multidimensional systemic analysis in urban morphological research and the data mining methods presented, over a univariate or bivariate analysis, are well documented. The initial selection of attributes, methods and tools used in the construction of the data set, represents an effort to condense the knowledge on public space from a comprehensive set of perspectives. The ongoing research suggests: (1) the application of a multidimensional method specifically designed for the urban square, an original contribution, as far as we know; (2) the integration of analyses at various scales; (3) the use of information covering the entire spectrum of spatial dimensions (0-1-2-3D); and (4) the synthesis of physical, spatial and immaterial attributes in a single descriptive vector. Despite the initial stage of attribute compilation, there is confidence in the validity of the approach, acknowledging that the analysis of the *corpus* aims to (re)define the method itself.

Correlations between attributes are not fully identified; however their inclusion in the initial set of attributes is important for completeness, even if later they are cast aside as redundant. Presently found limitations relate to the poor scalability of the parametric tools, some restraints in the cartographic databases and difficulty in completing information for some more remote settlements.

In future work the potential of the method points out four directions. First, to extend the examples to other types of contemporary urban spaces that challenge the traditional concept of square (large scale urban regeneration projects, commercial or cultural complexes, etc.), testing the generality of the method. Second, to study the receptivity of data mining in the urban morphology community, submitting the results to criticism. Third, to deepen the potential of data mining in urban analysis, exploring other dimensional reduction and learning techniques (supervised and unsupervised) less used or unexplored, exposing their potential and measuring the consistency of the initial results. Examples are: hierarchical clustering and *dendrograms*; decision trees and the modelling of the minimum set of association rules leading to the identified clusters; or the use of classification algorithms in determining correlations between clusters or examples and territorial or temporal distribution, or other labelled attributes. And finally, to explore the combination with generative methods of design.

ACKNOWLEDGMENTS

The authors thank the collaboration of FORMA URBIS Lab (FAUL), and professors Dias Coelho and Sérgio Fernandes. This work refers to an ongoing PhD research project funded by the FCT research grant SFRH/BD/95148/2013.

REFERENCES

Beirão, J, Chaszar, A and Cavic, L 2014 'Convex-and Solid-Void Models for Analysis and Classification of Public Spaces', *Proceedings of the 19th CAADRIA International Conference*, Kyoto, pp. 253-262

Benedikt, ML 1979, 'To take hold of space: isovists and isovist fields', *Environment and Planning B: Planning and Design*, 6(1), pp. 47-65

Campos, MB 1997 'Strategic spaces: patterns of use in public squares of the city of London', *Space Syntax - First International Symposium Proceedings*, London, pp. 26.1-26.11

Campos, AMB and Golka, T 2005 'Public Spaces Revised: of the relationship between Patterns of Stationary Activities and Visual fields', *Proceedings, 5th International Space Syntax Symposium*, Delft

Cataldi, G, Maffei, GL and Vaccaro, P 2002, 'Saverio Muratori and the Italian school of planning typology', *Urban Morphology*, 6(1), pp. 3-14

Chaszar, A and Beirão, J 2013 'Feature Recognition and Clustering for Urban Modelling-Exploration and Analysis in GIS and CAD', *Proceedings of the 18th Conference*, Singapore, pp. 601-610

Coelho, CD 2008, *A praça em Portugal. Inventário do espaço público*, DGOTDU, Lisbon

Cutini, V 2003 'Lines and squares: towards a configurational approach to the morphology of open spaces', *4th International Space Syntax Symposium,*, London

Frankhauser, P 1994, *La Fractalité des structures urbaines*, Anthropos, Paris

Gil, J, Beirão, JN, Montenegro, N and Duarte, JP 2012, 'On the discovery of urban typologies: Data mining the many dimensions of urban form', *Urban Morphology*, 16(1), pp. 27-40

Halford, GS, Baker, R, McCredden, JE and Bain, JD 2005, 'How many variables can humans process?', *Psychological Science*, 16(1), pp. 70-76

Hanna, S 2009 'Spectral comparison of large urban graphs', *Proceedings of the 7th International Space Syntax Symposium*, Stockholm

Hanzl, M 2013 'Modelling of Public Spaces', *Proceedings of the 31st eCAADe Conference*, Delft, pp. 319-327

Hillier, B 1996, *Space is the machine: a configurational theory of architecture*, Cambridge University Press, Cambridge

Hillier, B and Hanson, J 1984, *The Social Logic of Space*, Cambridge University Press, Cambridge

Laskari, A 2014 'Multidimensional Comparative Analysis for the Classification of Residual Urban Voids', *Proceedings of the 32nd eCAADe Conference*, Newcastle upon Tyne, pp. 283-292

Laskari, A, Hanna, S and Derix, C 2008, 'Urban Identity Through Quantifiable Spatial Attributes', *Design Computing and Cognition'08*, 2008 SRC, pp. 615-634

Berghauser Pont, M and Haupt, P 2010, *Spacematrix: Space, Density and Urban Form*, NAi Pub., Rotterdam

Psarra, S and Grajewski, T 2001 'Describing Shape and Shape Complexity Using Local Properties', *3th International Space Syntax Symposium*, Atlanta

Turner, a, Doxa, M, O'Sullivan, D and Penn, a 2001, 'From isovists to visibility graphs: A methodology for the analysis of architectural space', *Environment and Planning B: Planning and Design*, 28(1), pp. 103-121

Urhahn, G and Bobic, M 1994, *A pattern image: a typological tool for quality in urban planning*, Thoth Publishers, Bussum

Witten, IH and Frank, E 2005, *Data mining : practical machine learning tools and techniques*, Morgan Kaufman, Amsterdam; Boston, MA