DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5

CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM

Sinan Cetinkaya, Melih Basaraner

SI | EN

IZVLEČEK

Prepoznavanje in opisovanje prostorskih vzorcev je izrednega pomena pri kartografski generalizaciji, saj s tem zagotavljamo kar največje ohranjanje vzorcev v okviru omejitev merila. Poravnani objekti so eden od običajnih vzorcev na topografskih kartah in v bazah. Prepoznavajo se z ustreznimi Gestaltovimi faktorji, kot so bližina, podobnost, skupna orientacija in zveznost. Raziskava je osredotočena na vprašanje, kako določiti lastnosti poravnav objektov, ki so prepoznane s samodejnimi ali ročnimi metodami. Tu smo z Delaunayevo triangulacijo in regresijsko premico/ krivuljo določili nove mere, ki ustrezajo Gestaltovim faktorjem. Povezava med merami in Gestaltovimi načeli je prikazana z odločitvenim drevesom. Zaradi primerjave in razvrstitve poravnav objektov glede na kakovost je bila z vsoto vseh mer izračunana indeksna vrednost. Dodatno je bila izvedena nadzorovana klasifikacija z algoritmom C4.5, pri čemer smo dobili odločitveno drevo, ki smo ga uporabili za povezovanje razredov kakovosti z vrednostmi mer in samodejni pripis poravnave razredu kakovosti. Rezultati raziskave kažejo, da so predlagane mere primerne za predstavitev Gestaltovih faktorjev. S predlaganimi metodami bi lahko pospešili in olajšali postopek določevanja lastnosti poravnav objektov pri generalizaciji topografskih kart.

ABSTRACT

Detection and characterisation of spatial patterns is crucial for cartographic generalisation since it entails preserving the patterns as much as possible within scale limits. Building alignments are commonly confronted patterns in the topographic maps/databases. They are perceptually recognised in accordance with relevant Gestalt factors, namely proximity, similarity, common orientation and continuity. This study is concentrated on how to characterise building alignments detected by automated or manual methods. To this end, new measures based on Delaunay triangulation and regression line/curve are established to correspond to the Gestalt factors. The relationship between the measures and Gestalt principles has been illustrated with a decision tree. An index value was computed by total sum of measures' values to compare and order alignments from quality aspect. Additionally, a supervised classification was performed with C4.5 algorithm thus a decision tree was obtained to be able to both associate the quality categories with the measure values and automatically assign alignments into a quality class. The findings demonstrate that proposed measures are substantially effective for representing Gestalt factors. The proposed methods can potentially enhance and ease the characterisation of building alignments in topographic map generalisation.

KLJUČNE BESEDE

lastnosti poravnav objektov, odločitveno drevo, Gestaltovi faktorji, kartografska generalizacija

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 552 |

1 INTRODUCTION

A pattern implies a discernible coherent configuration based on the interrelationship between objects of interest through visio-cognitive processes or expert knowledge. When viewing the map, the eye discerns the spatial patterns – patterns of shape, orientation, connectedness, density and distribution. The user begins to record them, to explore and categorise these patterns in terms of the processes that formed them (Mackaness and Edwards, 2002). Spatial patterns are regular spatial organisations that can be perceived from spatial data sets. Steiniger (2007) distinguishes two kinds of spatial patterns- visual and geospatial. Visual patterns are the result of processes of perceptual organization without the use of domain knowledge, while geospatial patterns are patterns that are accessible only to persons with specific domain knowledge. In this context, alignments of buildings are identified as a visual pattern in conjunction with Gestalt factors (proximity, common orientation, similarity, continuity etc.; see Wertheimer (1923)) between the buildings.

Generalisation entails explicit modelling and extracting mostly implicit spatial patterns and relationships in order to keep or enhance them. It is critical to preserve patterns in generalisation because when cartographers generalise maps, they never restrict their view and analysis to the position of one object on the contrary they consider contextual relationships between objects and analyse them to be able to convey geographic information from source product to the target products (Ruas, 1998). A difficulty in generalisation is the need for data enrichment in order to identify and represent the implicit geographic phenomena before manipulating them. In order to keep the alignments during generalisation requires not only identify aligned buildings but also characterise them (Mustiere and van Smaalen, 2007). A characterisation is necessary to compare different structures to decide which one is the most structuring. To characterise the identified aligned structures, indicators are computed to define whether a structure is morphologically regular and/or extensionally important (Boffet and Rocca Serra 2001).

Related works have introduced the concepts and techniques for the detection, characterisation and generalisation of building alignments and a multi-scale behaviour of the patterns. Regnauld (2001) proposes a method of selection based on the typification principle that creates a result with fewer objects, but preserves the initial pattern of distribution. For this purpose, he uses a graph of proximity on the building set, which is analysed and segmented with respect to various criteria, taken from Gestalt theory. This analysis provides geographical information that is attached to each group of buildings. The information from the analysis stage is used to define methods to represent them at the target scale. The aim is to preserve the pattern as far as possible, preserve similarities and differences between the groups with regard to density, size and orientation of buildings. Boffet and Rocca Serra (2001) use three indicators to characterise the identified building alignments: 1) number of buildings, 2) homogeneity of the centroid distances, 3) size homogeneity. They state that the last one seems to be the most relevant one. Christophe and Ruas (2002) present a method to both detect and characterise building alignments to assist two contextual generalisation operations namely typification and displacement. The detected alignments are characterised with those perceptual criteria: proximity, arrangements, size, shape and orientation of the buildings. Mackaness and Edwards (2002) investigate the behaviour and evaluation of patterns at large changes in scale and those qualities that should be invariant at small

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 553 |

scales. Ruas and Holzapfel (2003) attempt to define the perceptual quality of building alignments. To this end, six parameters, i.e. alignment (continuity), distance, shape, size, orientation and stretching are identified and then they try to obtain similar results by adjusting weights of the parameters with the ones provided from the alignment rankings by expert cartographers. Li et al. (2004) give the factors having an impact on the recognition of building alignments based on Gestalt principles. After that, several geometric measures, such as the sum of the building area, mean separation and standard deviation of the separations, are assigned to each alignment. An appropriate operation is then selected to generalise a building alignment by means of the assigned information. Yan et al. (2008) focus on matching the characteristics of the generated building groups to appropriate generalization operations and algorithms with a series of rules based on parameters such as the number of buildings involved, the size of buildings, the ratio of building area and free space, etc. as well as threshold values such as a separation threshold and an area threshold. Zhang (2012) determines the homogeneity of building patterns based on the standard deviations of relevant properties, namely spacing (nearest distance), size, orientation, and shape with equal weights.

Characterisation of building alignments is relatively less researched area. Size, shape, orientation and inter-distance measures have been commonly used parameters for their characterisation. Of all these parameters, shape and orientation definitions are somewhat fuzzy and not enough representative. Besides they can negatively affect the characterisation process. Therefore, this article aims to: 1) discover new alternative measures to qualify the alignments, 2) illustrate the relationship between the measures and the Gestalt factors, 3) categorise the alignments based on the new measures with supervised classification. This article is organised as follows: in section 2, measures proposed in previous studies are described. New alternative measures and methods are introduced in section 3. Finally, experimental study is explained in section 4.

2 GESTALT PRINCIPLES AND COMMON MEASURES

2.1 Gestalt principles

Gestalt principles guide the study of how people perceive visual components, instead of many different parts, to formulate the regularities according to which the perceptual input is organized into unitary forms. Six main Gestalt factors determine how the visual system automatically groups elements into patterns: Proximity, Similarity, Closure, Symmetry, Common Fate, and Continuity. Common orientation factor is also added to the list with respect to specific cartographic patterns by Li et al. (2004) and Yan et al. (2008). For building alignments, closure is special case of continuity and symmetry is very rarely confronted factor and also rather complicated to measure so they have not been used in this study as well as in the related studies. The principle of common fate is only relevant in dynamic maps (Yan et al. 2008). Therefore, only the factors, given in Table 1 are taken into consideration for building grouping in this research.

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 554 |

Table 1:   Gestalt factors and their corresponding terms used in this study

Gestalt Factors	Description	Corresponding Terms
Similarity	elements tend to be integrated into groups if they are similar to each other	Size Shape
Common Orientation	elements arranged in a similar direction are perceived as a group	Orientation
Proximity	elements tend to be perceived as aggregated into groups if they are near each other	Inter-Distance Stretching
Continuity	oriented units or groups tend to be integrated into perceptual wholes if they are aligned with each other	Continuity

2.2 Common measures in the literature

In this section, descriptions of the measures for the commonly used Gestalt factors in the related works are given.

Size is measured with area for buildings in all of the previous works and simple standard deviation is computed for determine the homogeneity changing.

Shape is difficult to define with single parameter. Several measures may possibly be required depending on the complexity and characteristics of a polygon. However, previous alignment characterisation studies employ single parameter when quantifying shapes. These parameters are shown in Table 2 where b_i is i[th] the building in an alignment.

Table 2:   Shape measures

Measures	Definition (equation)	Source
Concavity	$\dfrac{Area\left(b_i\right)}{Area\left(convexHull(b_i)\right)}$	Ruas and Holzapfel (2003)
Compactness	$\dfrac{Perimeter\left(b_i\right)}{2*\sqrt{\pi*Area\left(b_i\right)}}$	Zhang (2012)
Edge number ratio	$\dfrac{Min\left(NumberOfEdges\left(b_i,b_j\right)\right)}{Max\left(NumberOfEdges\left(b_i,b_j\right)\right)}$	Yan et al. (2008)

In order to derive a shape homogeneity value, Ruas and Holzapfel (2003) just employ standard deviation while Zhang et al. (2013) use Equation 1. Yan et al. (2008) use edge number ratio to evaluate shape similarity between two adjacent buildings.

$$Shape\ homogeneity = 1 - STD_{(Si)}/Mean_{(Si)} \qquad (1)$$

where $STD_{(Si)}$ denotes standard deviation of shape values. $Mean_{(Si)}$ stands for average shape value.

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 555 |

Orientation measures for buildings are elaborately investigated by Duchene et al. (2003). Former studies propose the following orientation measures in the alignment characterisation (Table 3).

Table 3: Orientation measures

Definition	Description	Source
Main wall orientation	Orientation of the longest edge of a building polygon. (mod [90º])*.	Ruas and Holzapfel (2003)
Statistically weighted wall orientation	Average value of the orientations of each edge weighted by their lengths. (mod [180º]).	Zhang et al. (2013)
SMBR orientation	Orientation of the longest edge of Smallest Minimum Bounding Rectangle (SMBR) of a building polygon. (mod [180º]).	Yan et al. (2008)

* *mod* means modulo.

Ruas and Holzapfel (2003) use standard deviation as in the shape measure to produce the homogeneity value. Zhang et al. (2013) uses statistically weighted wall orientation and computes orientation homogeneity value according to Equation 2.

$$Orientation\ homogeneity = 1 - STD(O_i)/NFactor \tag{2}$$

where *NFactor* is a normalising factor and equals to 45°. *STD(O_i)* means standard deviation of orientations.

Inter-distance is defined with nearest distances except the study of Boffet and Rocca Serra (2001) in which centroid distances are used. Table 4 shows the summary of the previously proposed inter-distance measures.

Table 4: Inter-distance measures

Definition	Description	Source	Homogeneity
Minimum Distance	Minimum distance value between two building polygons	Ruas and Holzapfel (2003)	Standard deviation
Minimum Distance	Minimum distance value between two building polygons	Zhang et al. (2013)	The ratio of standard deviation to mean value
Centroid Distance	Centroid distance from each building to the all other buildings	Boffet and Rocca Serra (2001)	Standard deviation normalised by maximum distance
Minimum edge of true connection triangles	The length of shortest triangle edge between two buildings	Yan et al. (2008)	N/A

Stretching is computed by Ruas and Holzapfel (2003) as the ratio of average distance between successive buildings to square root of the average building area in the alignments.

Continuity called as alignment by Ruas and Holzapfel (2003) is measured through a regression line created using the centroids of buildings. Average distance from each centroid of a building to the regression line is used as a measure.

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 556 |

3 METHODOLOGY FOR THE CHARACTERISATION OF BUILDING ALIGNMENTS

First of all, we have to denote that this methodology needs building alignments to have been detected beforehand. In the methodology, new measures are proposed using auxiliary data structures based on Delaunay triangulation and regression line/curve. Five measures (area of free space –AoFS, triangle edge index –TEI, building density in the alignment –BDA, continuity –Cont., and alignment elongation –A. Elon.) that are the alternatives to the measures described in section 2 are introduced. Then using these measures, following issues are examined:

— relationship between proposed measures and Gestalt factors and the representative capability of the measures,
— derivation an index value for quantitative evaluation of each alignment and making a manual classification that is going to be used to obtain a decision tree in the next issue,
— qualitative evaluation of the alignments using C4.5 decision tree algorithm.

Figure 1 shows general flow chart of the proposed methodology for the characterisation of the building alignments.
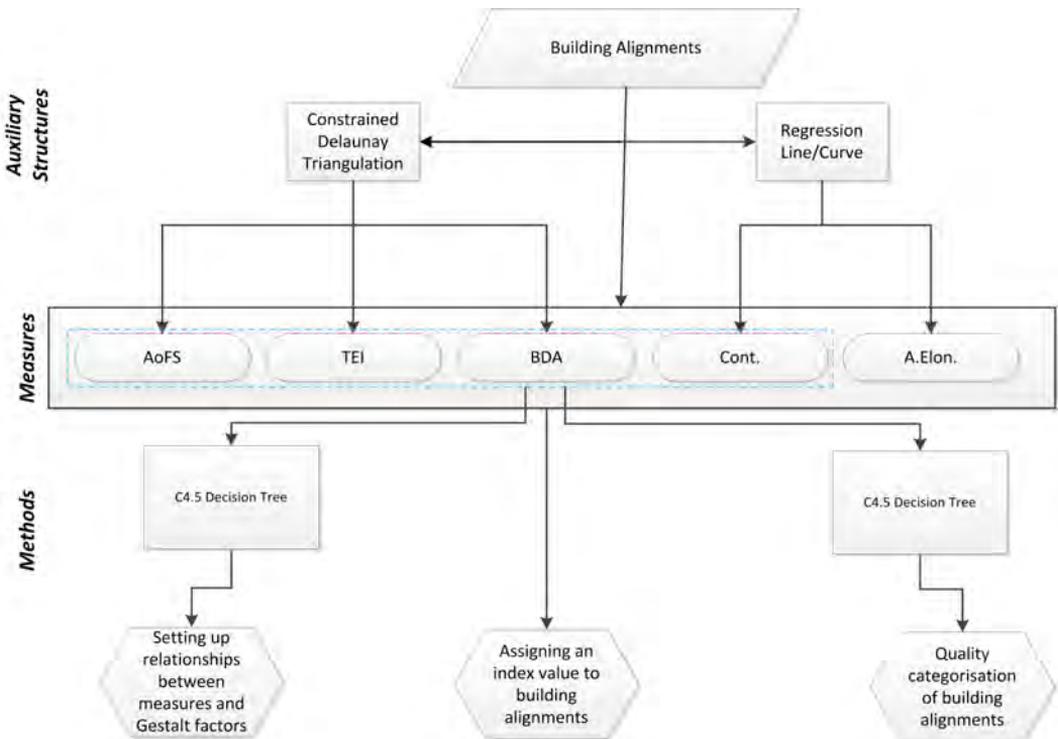

Figure 1:   Flow chart of the proposed methodology.

3.1 Measure definitions

Two kinds of measures are established for alignment characterisation: a) Free Space (FS)-based measures, b) the building vertex-based measures. These measures are the alternative measures to the ones presented in the Table 1 and have been analysed in Section 4.1 to reveal whether they represent the Gestalt factors.

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 557 |

In the remaining sections of the article, mean values and standard deviations are denoted by μ and σ in the following equations respectively.

3.1.1 Free Space-based measures

Three measures were defined based on FS: area of FS, edges of triangles in FS and building density in alignment. Equations of these measures were given in Equations 3, 4 and 5.

FS is the area between two buildings, formed by the triangles obtained through Constrained Delaunay Triangulation (CDT) (Figure 2). Triangulation process must obey the following rules:

— Triangle edges cannot intersect with buildings' edges.
— Buildings cannot contain any triangle.
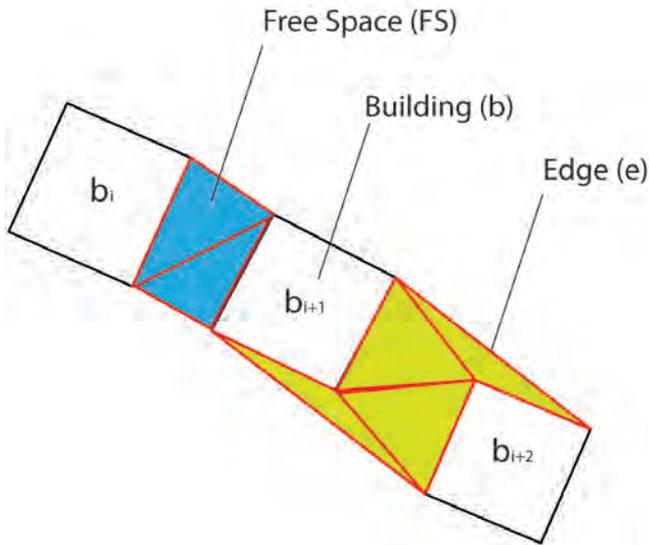— Triangles must connect two consecutive buildings.

Figure 2: Constrained Delaunay triangulation between successive buildings

Area of FS (AoFS) index is the ratio of the standard deviation to the mean value of FS areas in an alignment (Equation 3).

$$AoFSindex = \frac{\sigma_{Area(FS)}}{\mu_{Area(FS)}} \tag{3}$$

— *Triangle Edge Index (TEI)* is obtained in two steps; at first step, edge length deviation is computed via length of triangle edges in each FS then triangle edge index is calculated with the ratio of standard deviation to the mean value of length deviations of the edges.

$$EdgeLengthDeviation = \sigma_{Length(Edges)} \text{ (Standard deviation of edges in a FS)}$$

$$TEI = \frac{\sigma_{EdgeLenghtDeviations}}{\mu_{EdgeLenghtDeviations}} \tag{4}$$

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 558 |

— *Building Density in Alignment (BDA)* is the ratio of total area of buildings to sum of total area of FS and building areas.

$$BDA = \frac{\sum_{i=1}^{N} Area(b_i)}{\sum_{i=1}^{N} Area(b_i) + \sum_{i=1}^{N-1} Area(FS)} \tag{5}$$

where N is the number of buildings in an alignment.

3.1.2 Building vertex-based measures

The following two measures are based on regression line/curve of the building vertices in an alignment. Regression line/curve and the alignment (as a single object formed by group of the buildings) are overlaid to obtain end points of the regression line (Figure 3). If buildings in an alignment are located along the y axes, regression line estimation becomes problematic because the slope of regression line goes infinity. To solve this problem, alignment is rotated about 30 degree via Helmert transformation.

— *Alignment Elongation (A. Elon.)* is ratio of two times of the maximum deviation of vertices from the regression line to length of the regression line.

$$A.Elon. = \frac{2 * \max(r_i)}{lenght(regressionLine)} \tag{6}$$

where r_i means distance between vertex and adjusted line (Figure 3). In order to obtain the width of the alignment approximately, $2*\max(r_i)$ has been used. In fact $\max(r_i)$ value can probably be smaller at the other side of the regression line/curve. But it is considered as trivial.
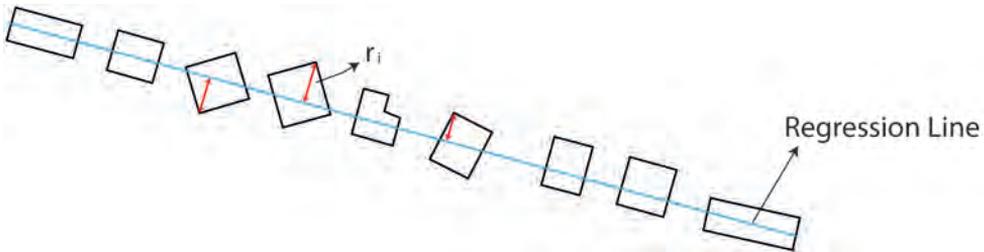

Figure 3: Regression line and perpendicular deviations (r_i).

— *Continuity* (*Cont.*) is the ratio of maximum deviation of building vertices to mean values of the building edges in an alignment.

$$Cont. = \frac{\max(r_i)}{\mu_{buildingEdgesInAlignment}} \tag{7}$$

3.2 Decision tree (C4.5)

Decision tree is a kind of supervised classification method used in several fields such as artificial intelligence and pattern recognition. It enables qualifying the building alignments easily if a simple decision tree can be established. Decision tree induction is the learning of decision trees from class-labelled training data. A decision tree is a flowchart-like tree structure, where each internal node (non-leaf node) denotes

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 559 |

a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label (Han et al. 2011).

C4.5 decision tree algorithm is introduced by Quinlan (1993). It uses gain ratio as splitting criteria. The splitting ceases when the number of instances to be split is below a certain threshold. Error–based pruning is performed after the growing phase. C4.5 can handle numeric attributes. It can induce from a training set that incorporates missing values by using corrected gain ratio criteria (Rokach and Maimon, 2005). In this study, two different types of the data are assumed to be suitable for generating decision trees: boolen (Section 4.1) and numeric (Section 4.3). Numeric attributes correspond to our proposed measures while the boolen data is derived from the thresholds manually determined for the measures based on the deterioration of each Gestalt factor. Result of the algorithm will generate threshold values for the each internal node (i.e. the measures).

4 EXPERIMENTS AND RESULTS

Twenty three alignments (from ID 1 to 23 in Table 5) have been selected from a topographic data set at the scale 1:10.000 for experimental testing. In addition, seven artificially created alignments (from ID 24 to 30 in Table 5; Figure 4) were included in test data set. For every alignment, five measures have been calculated (Table 5). To reveal the alignment characteristics, three tests have been carried out. First, in Section 4.1, the relationship between proposed measures and Gestalt factors has been investigated. Second, alignments have been ranked manually classified through the index values based on the total sum of the values of the measures in Section 4.2. Finally, in Section 4.3, the alignments have been automatically classified in five categories by a decision tree according to their index values.

Table 5: The values of the measures belonging to the alignments. Underlined and italic IDs correspond to the artificial alignments

ID	AoFS	TEI	BDA	Cont.	A.Elon.	ID	AoFS	TEI	BDA	Cont.	A.Elon.
1	0.652	0.319	0.576	1.635	0.135	16	0.444	0.875	0.323	0.787	0.069
2	0.169	0.183	0.362	0.704	0.124	17	0.183	0.424	0.407	0.654	0.104
3	0.256	0.331	0.506	1.157	0.091	18	0.404	0.508	0.752	2.205	0.128
4	0.287	1.157	0.130	0.327	0.050	19	0.359	0.291	0.663	1.596	0.117
5	0.410	0.532	0.350	1.196	0.137	20	0.292	0.213	0.638	2.067	0.202
6	0.077	0.119	0.558	0.864	0.084	21	0.477	0.196	0.352	1.566	0.192
7	0.294	0.641	0.444	1.089	0.095	22	0.376	0.332	0.532	1.777	0.249
8	0.083	0.015	0.314	0.839	0.159	23	0.222	0.171	0.760	2.387	0.118
9	0.512	0.555	0.465	0.944	0.101	_24_	0.000	0.000	0.224	0.576	0.090
10	0.499	0.827	0.272	0.640	0.070	_25_	0.098	0.034	0.298	1.168	0.138
11	0.234	0.160	0.167	0.294	0.078	_26_	0.111	0.160	0.473	0.991	0.148
12	0.386	0.531	0.407	1.129	0.177	_27_	0.342	0.395	0.244	0.763	0.122
13	0.619	0.361	0.312	0.533	0.168	_28_	0.652	0.742	0.437	0.523	0.059
14	0.325	0.554	0.448	0.777	0.089	_29_	0.398	0.110	0.585	3.213	0.461
15	0.487	0.504	0.405	0.841	0.180	_30_	0.000	0.000	0.414	0.519	0.061

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 560 |

4.1 Revealing the relationship between measures and Gestalt factors

The artificial alignments have been utilised to detect how changes in alignments with respect to each Gestalt factor affect the measures. In this respect, one of them (ID 24) was designed as an ideal alignment (i.e. a nearly perfect alignment in view of all Gestalt factors) and only one Gestalt factor was deteriorated in each scenario (Figure 4).
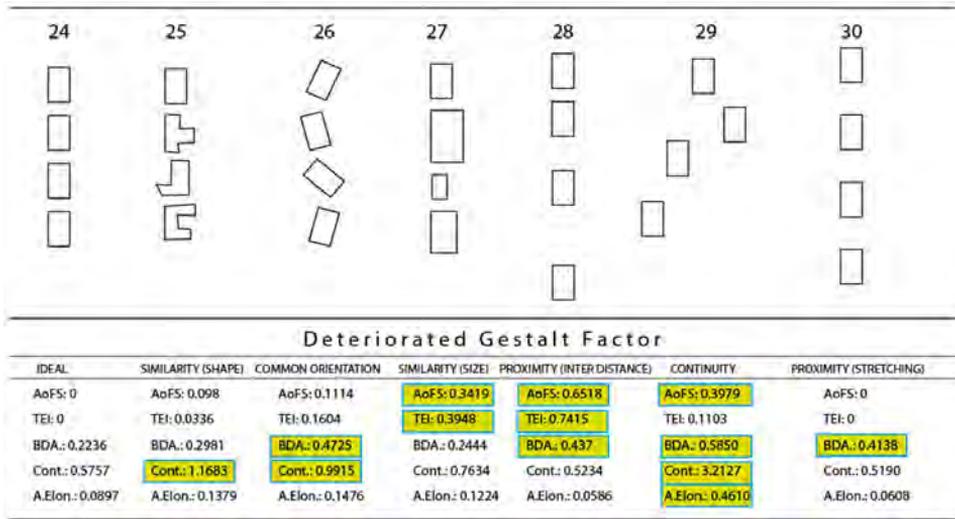

Figure 4: Artificial alignments for detection of relation between measures and Gestalt factors. The green boxes denote the substantial changes on the measures during the deterioration of a Gestalt factor.

Deviation of each measure from ideal case was investigated based on Gestalt factors and it was revealed which measure dramatically changes in which case. For example, it can be seen in Figure 4 that building density in alignment (BDA) and continuity (Cont.) measures change owing to the deterioration of common orientation. Experimentally determined threshold values were used to find out whether any measure influences any Gestalt factor. Accordingly, if value of a measure exceeds its threshold value then Boolean value of the measure pertaining to Gestalt factor is assigned 'True' otherwise 'False' (Table 6).

Table 6: Measures and Gestalt factors; numeric values under the captions denote assigned thresholds. This table was obtained by using Figure 4. Green boxes in Figure 4 correspond to the 'True' value.

Measures					Gestalt factor
AoFS >= 0.30	TEI >= 0.30	BDA >= 0.41	Cont. >= 0.97	A.Elon. >= 0.28	
False	False	False	True	False	Similarity (Shape)
False	False	True	True	False	Common Orientation
True	True	False	False	False	Similarity (Size)
True	True	True	False	False	Proximity (Inter-Dist.)
True	False	True	True	True	Continuity
False	False	True	False	False	Proximity (Stretching)

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 561 |

Figure 5 shows the decision tree obtained by employing C4.5 algorithm with the input values given in Table 6 using WEKA™ software (Witten et al. 2011). This decision tree enables to reveal the most influential Gestalt factor that reduces the alignment quality. Figure 5 also illustrates an example of the determination process of the most deteriorating Gestalt factor. Using the calculated measure values of an alignment, determination of the most deteriorated Gestalt factor is accomplished by progressing through the decision tree step by step.
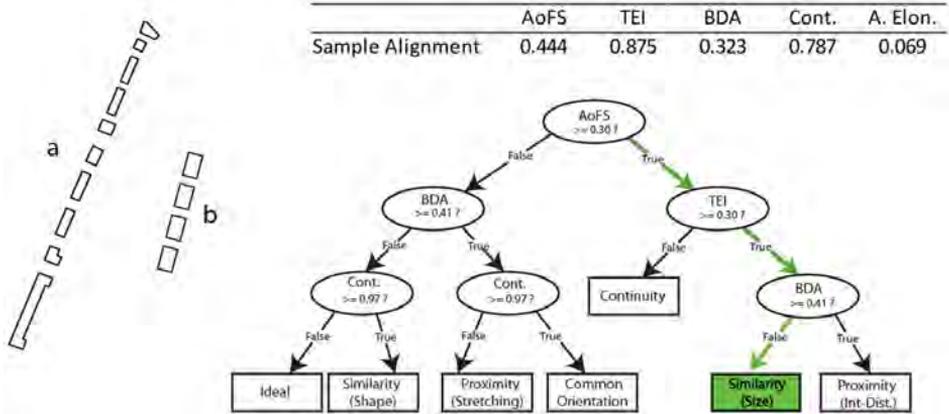
Figure 5: The obtained decision tree and a sample alignment illustrating the relations between the measures and Gestalt factors. This decision tree shows that the most deteriorated Gestalt factor of the sample alignment (a) is "Similarity (Size)" in comparison to the ideal alignment (b).

4.2 Generating index values to compare alignments and supervised categorisation for the classification

An index value has been calculated with total sum of the measure values for each alignment. The smaller the index value is, the better the quality of an alignment is. So this index value can be useful in order to compare quality of alignments quantitatively.

Sorted alignment index values in ascending order have been given in Figure 6. A classification has been manually performed based on leaps in the index values shown on the chart. Five classes have been distinguished for qualifying the alignments: *Very Good*, *Good*, *Average*, *Bad*, and *Very Bad*.
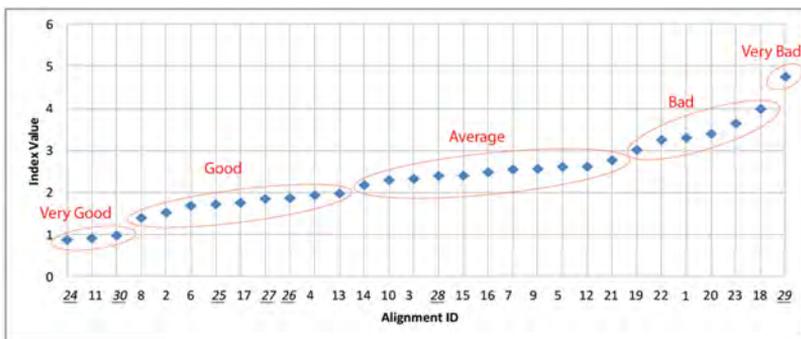

Figure 6: Chart of sorted index values of the alignments. Artificial ones are shown in italic and underlined style.

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 562 |

Results of both quantitative and qualitative assessment of the alignments can be seen in Figure 7. In the next section, a classification is made via decision tree by using the manually assigned class values.
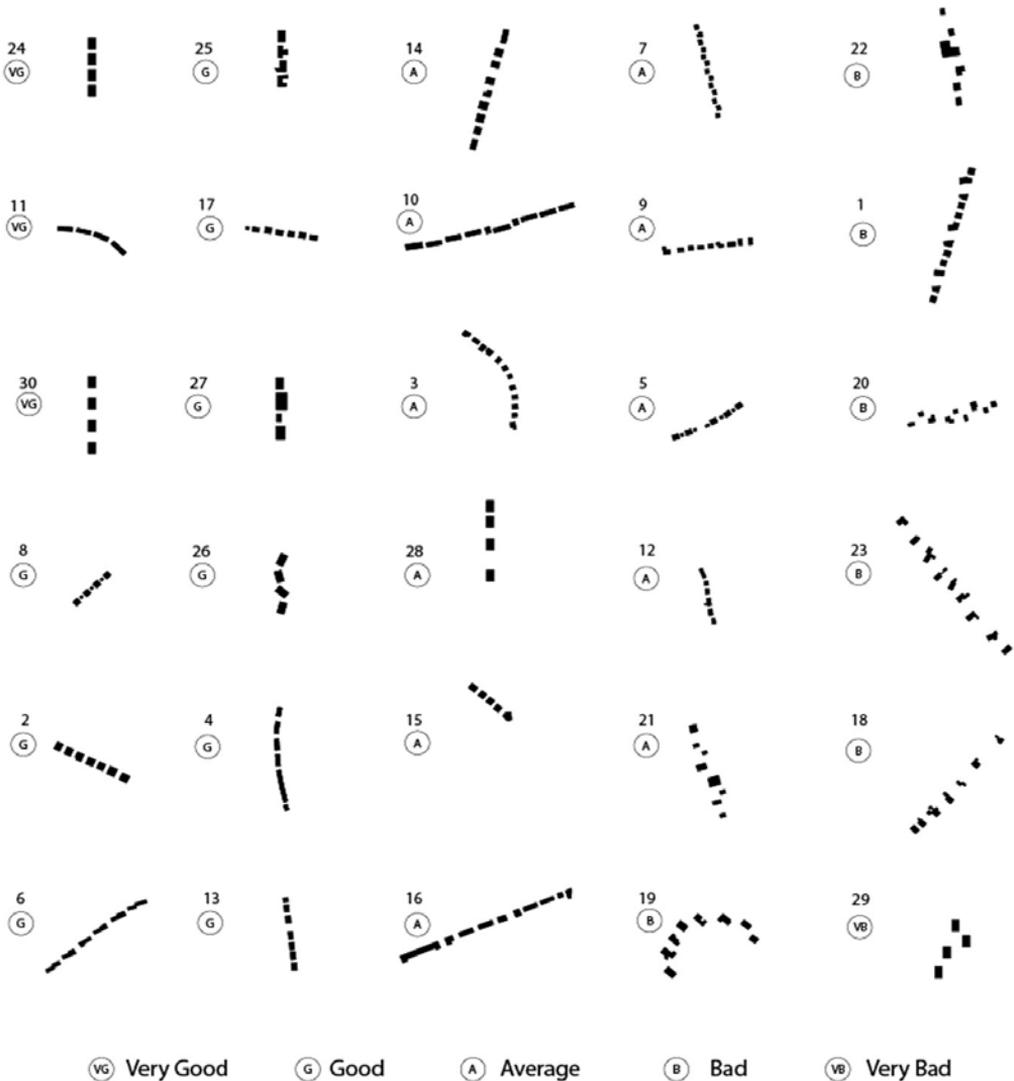

Figure 7: Alignments ordered by indexed values and their assigned quality classes

4.3 General classification of alignments by decision tree

A classification process is required in order to qualitatively categorise the alignments. For this purpose, a decision tree has been derived through C4.5 algorithm in WEKA™ by employing the measure values and the assigned quality classes as input data. The obtained classification tree is demonstrated in Figure 8.

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |
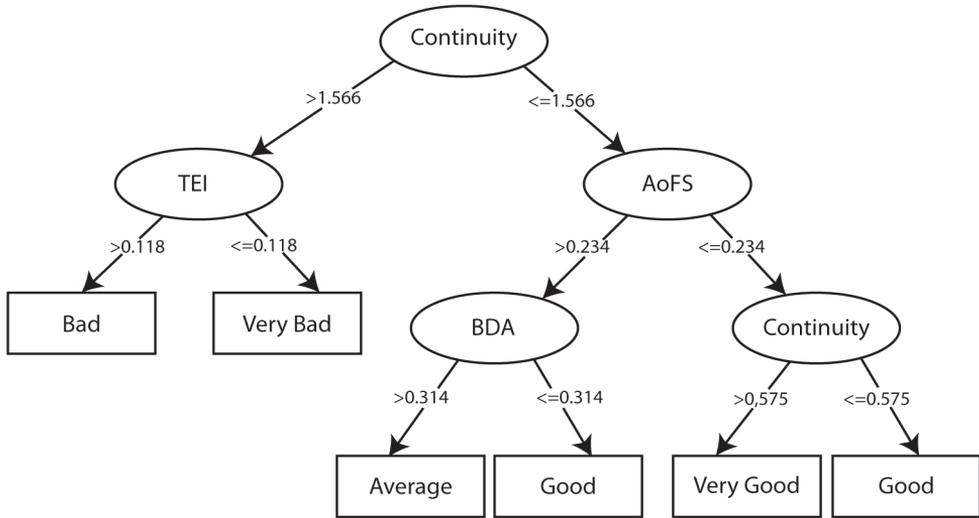
| 563 |

Figure 8:   Decision tree for qualifying building alignments.

The confusion matrix of the decision tree is given in Table 7. The success rate of fitting the values with class label into the decision tree is 96.7%. Just one alignment has been wrongly classified as good instead of average. Thus, an alignment can be qualified based on the measures through this decision tree.

Table 7:   Confusion Matrix

		Predicted Class				
		Very Good	Good	Average	Bad	Very Bad
	Very Good	3	0	0	0	0
	Good	0	9	0	0	0
Actual Class	Average	0	1	10	0	0
	Bad	0	0	0	6	0
	Very Bad	0	0	0	0	1

5 DISCUSSIONS

It has been proven by the experimental testing that proposed five alternative measures can reflect the changes in the Gestalt factors. So we can say that our measures are valid and relevant for the alignment characterisation. Sum of the measures has been used to get the one index value for the quantitatively comparison of the alignments and to manually classify the alignments in five groups. Finally a decision tree has been established to assign an alignment a quality class by using manually qualified alignments. In other words, this decision tree enables which alignment is better perceived graphically according to Gestalt factors. Therefore, this information guide the contextual generalisation of buildings to communicate this kind of spatial pattern more correctly at smaller scales (Basaraner and Selcuk, 2008). If one needs to know about the quality order of alignments, an index value can be calculated by the sum of the measures (i.e. smaller values correspond to better quality). Furthermore, proposed characterisation methods can be embedded into detection of the building alignments in an iterative manner. In other

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 564 |

words, during the detection of the alignments, alignment candidates can be compared to each other to select the best one.

When compared to the studies in the literature (Ruas and Holzapfel 2003; Zhang 2012), not only alternative measures have been proposed but also a simple decision tree has been established to assign a quality class to an alignment. In comparison to the previously employed orientation and shape measures, our measures offer a powerful alternative with respect to Gestalt factors. For example, continuity measure proposed by Ruas and Holzapfel (2003) does not consider the size of buildings (scale factor) because it uses the mean deviation value. But continuity measure proposed in this study is scale independent because it is an index value (ratio of maximum deviation to the average building edge).

All the proposed measures in this study are scale independent because all of them are proportional values, hence they can be used at all relevant scales. The results show the usability of the proposed measures while automated calculation of them especially based on the free space is non-trivial process. Decision tree thresholds can be said generic enough since our data set includes variety of building alignments. Although, by using the decision tree, an alignment can be easily assigned one of five classes in maximum three query steps (see Figure 8), precision of the decision tree can be improved by incorporating more alignment samples. Alignment elongation measure has not found a place in the decision tree because the other four measures are sufficient to discriminate the alignment data set.

Regression curve may not be fitted well for curvilinear alignments owing to the lack of mathematical constraints. In this case, elongation and continuity measures are negatively affected. Therefore, regression curves should be automatically drawn and visually checked beforehand (Figure 9). We used minimum area bounding rectangle during the construction of the Delaunay triangulation. This approach may have little effect on the values of the measures based on free space.
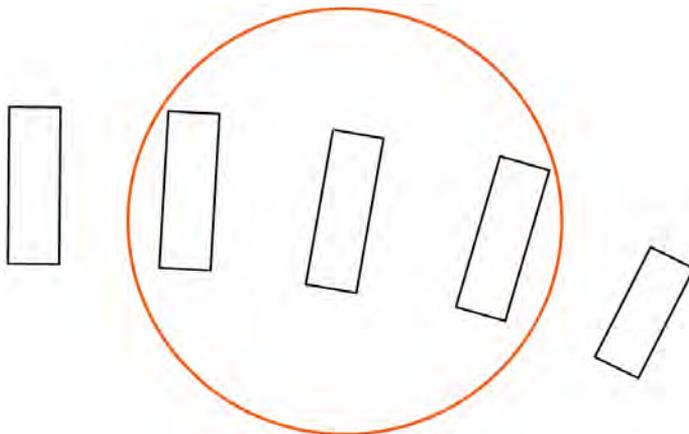

Figure 9: Circle regression problem

6 CONCLUSIONS

This paper has proposed new measures and approaches for building alignment characterisation. Five measures have been developed that corresponds to Gestalt factors. Three of them are derived from free

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 565 |

spaces constructed through constrained Delaunay triangulation. Two of them are based on the regression line/curve. Three implementations have been performed regarding to measures, indexing and classification.

First of all, alternative measures have been developed to represent the relevant Gestalt factors. Relationships between each factor and measures have been established via a decision tree trained with artificial alignment data. Second, quality index values of the alignments have been computed with total sum value of the all measures for comparison. Five quality categories have been manually determined by means of the sorted index chart and then the categories have been assigned to the alignments as a class attribute. Finally, a decision tree has been obtained in order to produce qualification criteria for alignments without quality category. It is easy to determine an alignment quality by means of the resulting decision tree. In contextual generalisation, selection of the generalisation operators will be more precise by using the proposed approaches.

As a future work, building alignment generalisation based on quality value can be investigated. In other words, relation between generalisation operators and building alignment quality can be examined with respect to the scale transitions.

Acknowledgements

References:

Basaraner, M., Selcuk, M. (2008). A structure recognition technique in contextual generalisation of buildings and built-up areas. Cartographic journal, 45 (4), 274–285. DOI: http://dx.doi.org/10.1179/174327708X347773

Boffet, A., Rocca Serra, S. (2001). Identification of spatial structures within urban blocks for town characterisation. Proceedings of 20th international cartographic conference (pp. 1974–1983), Beijing.

Christophe, S., Ruas, A. (2002). Detecting building alignments for generalization purposes. D.E. Richardson and P. van Oosterom (eds.), Advances in spatial data handling (pp. 419-432), Berlin: Springer.

Duchêne, C., Bard, S., Barillot, X., Ruas. A., Trévisan, J., Holzapfel, F. (2003). Quantitative and qualitative description of building orientation. 7th ICA workshop on progress in automated map generalisation, 28-30 April, Paris.

Han, J., Kamber, M., Pei, J. (2011). Data Mining: Concepts and Techniques. 3rd edition. Waltham: Morgan Kaufmann.

Li, Z., Yan, H., Ai, T., Chen, J. (2004). Automated building generalization based on urban morphology and Gestalt theory. International journal of geographical information science, 18 (5), 513–534. DOI: http://dx.doi.org/10.1080/13658810410001702021

Mackaness, W. A., Edwards G. (2002). The importance of modelling pattern and structure in automated map generalisation. 6th ICA workshop on progress in automated map generalization, 7-8 July, Ottawa, Canada. http://www.ikg.uni-hannover.de/isprs/workshop/macedwards.pdf

Mustière, S., van Smaalen, J. (2007). Database requirements for generalisation and multiple representations. W.A. Mackaness, A. Ruas and L.T. Sarjakoski (Eds.), Generalisation of geographic information: cartographic modelling and applications (pp. 113–136). Amsterdam: Elsevier.

Quinlan, R., (1993). C4.5: Programs for Machine Learning. San Mateo: Morgan Kaufman.

Regnauld, N. (2001). Contextual building typification in automated map generalization. Algorithmica, 30, 312–333. DOI: http://dx.doi.org/10.1007/s00453-001-0008-8

Rokach, L., Maimon, O. (2005). Decision trees. The Data Mining and Knowledge Discovery Handbook, (pp. 165–192). New York: Springer.

Ruas, A. (1998). O-O constraints modelling to automate urban generalisation process. Proceedings of 8th international symposium on spatial data handling (pp. 225–235). Vancouver, Canada.

Ruas, A., Holzapfel, F. (2003). Automatic Characterisation of Building Alignments By Means of Expert Knowledge. Proceedings of the 21st international cartographic conference, 10–16 August 2003, Durban, South Africa.

Steiniger, S. (2007). Enabling pattern-aware automated map generalization. PhD thesis. Zurich: Faculty of science, University of Zurich.

Wertheimer, M. (1923) Laws of organization in perceptual forms. In: Ellis WD (ed) A source book of gestalt psychology. Routledge & Kegan Paul, London, pp 71–88.

Witten, I. H., Frank, E., Hall, M. A. (2011) Data Mining: Practical Machine Learning Tools and Techniques. 3rd edition, Waltham: Morgan Kaufmann.

| 566 |

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

Yan, H., Weibel, R., Yang, B., (2008). A multi-parameter approach to automated building grouping and generalization. Geoinformatica, 12 (1), 73–89. DOI: http://dx.doi.org/10.1007/s10707-007-0020-5

Zhang, X. (2012). Automated evaluation of generalized topographic maps. PhD thesis. Enschede: Faculty of geo-information science and earth observation, University of Twente.

Zhang, X., Stoter, J., Ai, T., Kraak, M.J., Molenaar, M. (2013). Automated evaluation of building alignments in generalized maps. International journal of geographical information science, 27 (8), 1550–1571. DOI: http://dx.doi.org/10.1080/13658816.2012.758264

M.Sc. Sinan Cetinkaya
Yildiz Technical University (YTU), Faculty of Civil Engineering, Department of Geomatic Engineering
34220 Esenler/Istanbul, Turkey
e-mail: sicetin@yildiz.edu.tr

Assoc. Prof. Dr. Melih Basaraner
Yildiz Technical University (YTU), Faculty of Civil Engineering, Department of Geomatic Engineering
34220 Esenler/Istanbul, Turkey
e-mail: mbasaran@yildiz.edu.tr

Sinan Cetinkaya, Melih Basaraner | DOLOČANJE LASTNOSTI PORAVNAV OBJEKTOV NA PODLAGI NOVIH MER Z UPORABO ALGORITMA ODLOČITVENIH DREVES C4.5 | CHARACTERISATION OF BUILDING ALIGNMENTS WITH NEW MEASURES USING C4.5 DECISION TREE ALGORITHM | 552-567 |

| 567 |