CHARACTERIZATION AND DETECTION OF BUILDING PATTERNS IN CARTOGRAPHIC DATA: TWO ALGORITHMS

Xiang Zhang [a,b,*] **Tinghua Ai** [b] **, Jantien Stoter** [c,d]

[a]ITCt University of Twente,he Netherlands
xzhang@itc.nl
[b]School of Resource and Environment Sciences, Wuhan University, China
tinghua_ai@tom.com
[c]Delft University of Technology, the Netherlands
[d]Kadaster, Apeldoorn, the Netherlands
j.e.stoter@tudelft.nl

KEY WORDS: Pattern Recognition, Building Pattern, Map Generalization, Delaunay Triangulation, Minimum Spanning Tree, Algorithm, Graph Theory

ABSTRACT:

Building patterns are important features in applications like automated generalization and spatial data mining. Many previous work has however focused on a few specific patterns (i.e. collinear pattern), while many others are less discussed. This paper proposes a comprehensive typology of available building patterns through the study of existing maps, and discusses their characteristics. This typology includes collinear, curvilinear, align-along-road, grid-like and unstructured patterns. Two algorithms are presented to detect align-along-road and unstructured building patterns, which are tested against a topographic dataset of the Netherlands.

1 INTRODUCTION

Building patterns are important features in urban and rural areas. The automated detection of visually significant building patterns is required for applications like automated map generalization, automated evaluation of generalized output, semantic enrichment of spatial databases, and spatial data mining. For example, collinear patterns have been extensively investigated (Boffet and Rocca Serra, 2001, Christophe and Ruas, 2002) in order to simplify and typify building groups. An approach to detect higher-level semantics like terraced house (Lüscher et al., 2009) made use of detected building alignments; the detection approach is however not widely applicable as the buildings in their case was topologically adjacent, which is not commonly the case. Therefore, a more comprehensive view of which building patterns are available, and a generic approach to detect and characterize them are required.

As for the detection techniques, a remarkable and comprehensive investigation has been made on the use of minimum spanning trees (MST) in the field of pattern recognition (Zahn, 1971). After being successfully applied to some classical clustering problems for point sets, this technique was applied to detect building clusters for generalization purposes (Regnauld, 1996). Nevertheless, no interesting building patterns is recognized with this technique except some general tree-like clusters.

This paper firstly proposes a typology of building patterns by mainly studying existing topographic maps (Section 2). Then Section 3 develops a Graph-theoretic based approach to the detection and characterization of two common patterns of the typology. The proposed algorithms are implemented and tested in Section 4. This paper ends with conclusions (Section 5).

2 TYPOLOGY AND CHARACTERISTICS OF BUILDING PATTERNS

A typology of building patterns is needed as it formalizes our knowledge on building structures available in geospatial domain. The typology of all building patterns that occurred in the studied

maps and previous literature are identified and characterized in Section 2.1, and then we focus specifically on align-along-road pattern, discussing its relationships to other linear patterns (Section 2.2).

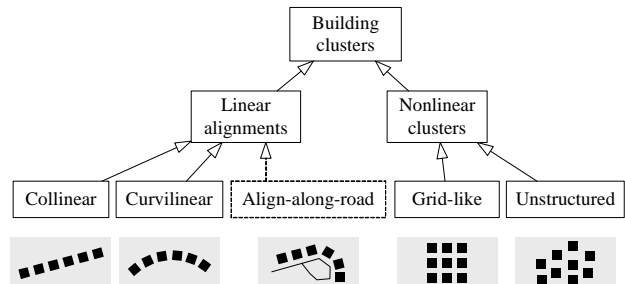### 2.1 Identifying and characterizing the typology

Figure 1: Typology and schematic examples of building patterns

The building patterns discussed in this paper are categorized as low-level, localized visual patterns, because they are important considerations in map generalization. The typology (Figure 1) is structured as follows. We define building patterns at top level as building clusters of spatially proximate objects with similar geometrical (e.g. spacing, size, orientation and shape) and semantic properties, extended from the definition of point patterns (Zahn, 1971). At an intermediate level, we refine the building clusters into linear alignments and nonlinear clusters in terms of 'group shape'. In general, the linear alignments appear to be more elongated and their constituent buildings can be organized by a linear path, while the nonlinear ones appear to stretch in two dimensions. At a finer level, the linear alignments are subdivided into collinear, curvilinear, and align-along-road patterns; the nonlinear clusters consist of grid-like and unstructured patterns.

We identify this typology from several sources. First, Gestalt principle of visual perception (Wertheimer, 1923) is used to define building patterns in general (proximity for general clusters) and specific (e.g. good continuity for linear alignments). As a

---

*Corresponding author. Email: xiangzhangchina@gmail.com

result, this typology is generic in the sense that all visual building patterns are clusters. Second, by comparing with those proposed by others we find that this typology is more comprehensive and generic. As mentioned before, collinear alignment has been acknowledged in the literature; grid-like pattern, though being less studied, has been discussed once for typification (Anders, 2006). He (Anders, 2006) also proposed a typology consisting of linear, circular, grid, star and irregular patterns. His linear pattern is enriched by our collinear and curvilinear alignments; meanwhile, his circular is a special case of curvilinear pattern and the star pattern are a combination of two crossing collinear alignments. We propose a further type (i.e. align-along-road) which integrates the relationship to surrounding roads. The typology proposed in this paper is regarded to be generic as all of the pattern types are confirmed by studying the maps (of the Netherlands, France, and Spain) in the EuroSDR generalization study (Stoter et al., 2009a), and maps of China, examples are shown in Figure 2. Other higher-level or global pattern can be seen as spatial combination of these low-level patterns and the integration of these patterns with other information such as semantics.

The typology is characterized as follows. According to previous work (Boffet and Rocca Serra, 2001, Christophe and Ruas, 2002, Ruas and Holzapfel, 2003), the homogeneities of general clusters can be realized using standard deviations of all the properties. In our approach, we improve this calculation by adopting the concept of coefficient of variance ($CV = Std/Mean$):

$$Homogeneity(P) = \frac{\sqrt{(p_i - Mean(P))^2}}{Mean(P)} \ , \qquad (1)$$

where P represents the properties of spacing, size, orientation, shape, and semantics of the cluster; $p_i \in P$ denotes the values measured from or between the cluster's elements.

The use of CV applies to the properties of spacing, size, shape and semantics. This is not because we want to normalize the properties but rather because CV is a dimensionless number. It means that the homogeneities become then relative numbers invariant to the choice of measurement, which is also consistent with our perception about building patterns. For example, if two patterns have the same standard deviation of size, the one with larger mean size (i.e. smaller CV) is more homogeneous in terms of size property. However, CV cannot be applied to orientation as it is meaningless to calculate CV for a cyclic variable. Therefore, the homogeneity of orientation is calculated using standard deviation.

The homogeneities (Equation 1) are common characteristics shared by all types of building patterns. In addition, most of the specific types of building patterns have their own characteristics. In the case of collinear alignments, the patterns are characterized by *straightness* describing the sinuosity degrees of the paths and *main angle* describing the directions of the paths. For curvilinear alignments, *smoothness* of the paths and their *curvature descriptions* should be emphasized. Align-along-road patterns should be attached to the roads along which they are aligned. Besides, it has an extra homogeneity, that is the homogeneity of distances to the aligned road. This property reflects the degree to which the pattern are parallel to the road. For the two nonlinear clusters, unstructured clusters have no specific property while grid patterns can be further characterized by *squareness* and *parallelism*. If we connect proximate buildings in a grid pattern together, we should get two sets of parallel lines, which intersect each other approximately right-angled. All patterns of the typology are illustrated using existing maps (Figure 2(a)-2(d)).
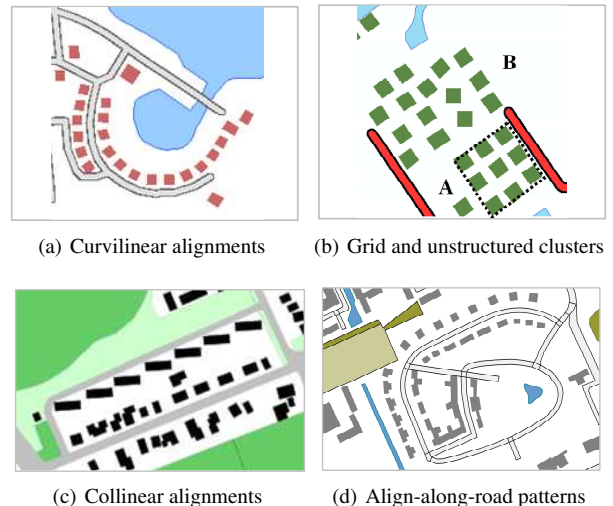

(a) Curvilinear alignments   (b) Grid and unstructured clusters

(c) Collinear alignments   (d) Align-along-road patterns

Figure 2: Examples of different building patterns as a result of studying existing mapsr

2.2 Align-along-road pattern and its relationship to other linear alignments

As a result of studying existing topographic maps, we find that the align-along-road patterns are one of the most common features which are visually significant in urban and rural structures; we also find that the distinction between align-along-road patterns and another two linear (i.e. collinear and curvilinear) alignments is not always clear. On the one hand, it is common cases that buildings are located near roads and streets, and therefore such collinear and curvilinear alignments (as shown in Figure 2(c) and 2(a)) are also align-along-road patterns. On the other hand, collinear and curvilinear alignments may be parts of align-along-road patterns, because normally align-along-road patterns are not regular in terms of curvatures. This sometimes means that a align-along-road pattern can be segmented into pieces of collinear and/or curvilinear alignments. Of course, there are also situations where collinear and curvilinear patterns are independent of roads.

We therefore limit ourselves to the detection and characterization of two patterns, i.e., the align-along-road and unstructured patterns. Some of the other patterns have been discussed by previous authors. For example, the collinear pattern by (Christophe and Ruas, 2002) and the grid pattern by (Anders, 2006).

3 DETECTION AND AUTOMATIC CHARACTERIZATION METHOD

In this section, we propose two Graph-theoretic based detection algorithms for align-along-road and unstructured patterns. Currently, semantics of buildings (e.g. detached/terraced house) is not available in most topographic datasets (Stoter et al., 2009b), and the semantic information is thus not discussed in this paper. It is nevertheless possible to integrate such information to refine detection results in future work.

This section first introduces the preliminary work which will be used by the subsequent detection and characterization (Section 3.1). Then the detection and automatic characterization for align-along-road (Section 3.2) and unstructured (Section 3.3) patterns are presented.

3.1 Basic computational tools

Refined constrained Delaunay triangulation   The constrained Delaunay triangulation (CDT) plays an important role in the following detection and characterization, and hence some fundamental computations based on the CDT are introduced at first.
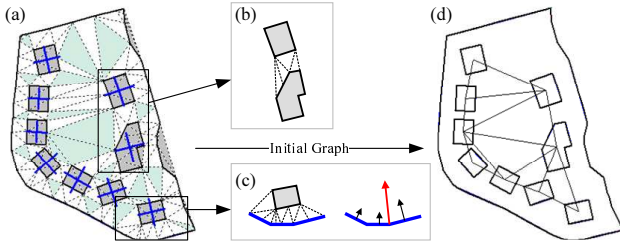
Figure 3: (a) Constrained Delaunay Triangulation takes buildings (grey polygons) and roads (bold lines) as constrained objects; (b) distances defined on the incident triangles between proximate objects; (c) segments of a road along which the building aligns and an illustration of the normal direction of this part of the road (the longest red arrow); (d) the initial proximity graph from (a)

As shown in Fig. 3(a), the CDT is built on buildings and roads, taking their outlines as constrained lines. The CDT is refined by inserting extra points to the constrained lines, and the interval between inserted points is based on the minimal distance between all data points. Two objects are considered as neighbors only when they are connected by edges of the triangles; the proximity relationship between two buildings, and between buildings and roads are explicitly modeled by this structure.

Based on the proximity relationship between buildings, an initial graph (Fig. 3(d)) can be generated from which the Minimum Spanning Tree (MST) can be automatically derived (see next section). Note that in the initial graph, and in the derived MST as well, all edges is actually weighted based on the proximity between building outlines, although for graphic presentations the edges connecting the centroids of buildings (i.e. vertices of the graph) are delineated. This means that the weights stored in the edges are calculated by the nearest distances between building outlines, instead of building centroids.

Several benefits can be drawn from the above CDT. Although being less efficient, it enables a better representation of proximity relationship than building a CDT on the centroids of buildings because it takes the shapes of buildings into account. In addition, because road centerlines are also involved in the CDT, the initial graph is automatically segmented in the sense that no graph edge intersects any dead end within a partition formed by the roads (Fig. 4).

As for the computational efficiency, the theoretical efficiency for the whole detection procedure (including constructing initial graph, deriving MST, detecting and characterizing building patterns) can be compensated for by the refined CDT. For example, the calculation of the nearest distance between two proximate objects can make use of their incident triangles (white triangles in Fig. 3(a), (b) and (c)), enabling a faster computation. That is, for each incident triangle between two objects, a distance (d_i) can be computed from the triangle edge that coincides with one of the objects to the third vertex; the nearest distance between the two objects is the minimum amongst all distances ($Min(\{d_i\})$). The time complexity of this calculation is $O(t)$, where t is the number of incident triangles between two objects; while a trivial computation of the nearest distance between two buildings (as was used in Regnauld, 1996) takes $O(nm)$ time, where n and m are the numbers of the points of the two buildings respectively. This speeds up greatly the nearest distance calculation between two buildings (Fig. 3(b)) and especially between a building and a nearby road (Fig. 3(c)), since t is generally much smaller than nm that can be expected for two spatial objects.

Other required calculations are as follows. First, the wall statistical weighting method (Duchêne et al., 2003) was implemented for computing building orientation; they showed that this
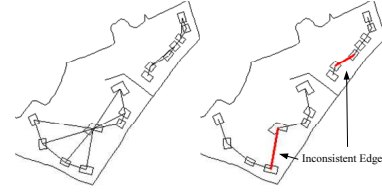

Figure 4: Initial graph, its derived MST and inconsistent edges

method suits well for describing the orientations of typical man-made features (see also the crosses in the buildings in Fig. 3(a)). Second, the normal direction of a portion of a road (the longest arrow in Fig. 3(c)) is computed based on the average of all the segments normal vectors (three shorter arrows in Fig. 3(c)) weighted by segment lengths. The two calculations are crucial for the detection of align-along-road patterns. The last calculation is shape index (AGENT, 1999), which is required to characterize the homogeneity of shape.

Minimum Spanning Trees In graph theory, a *spanning tree* of an undirected graph G is a tree that contains all vertices of G; the weight of a tree is defined as the sum of the weights of all its constituent edges. A *minimum spanning tree* of G is then a spanning tree whose weight is the minimum among all spanning trees of G. Since a graph may not necessarily be connected (Fig. 3a), it has a union of minimum spanning trees for its connected components (Fig. 3b). In this work, the vertices of G represent building features; the edges record their proximity relationships and are weighted by the nearest distances. Prim's algorithm (Prim, 1957) which is more efficient than Kruskal's algorithm (Zahn, 1971) was implemented to derive MST from initial graph.

A core concept for MST in pattern recognition is so-called inconsistent edge proposed by Zahn (1971), where he provided evidence that MST and inconsistent edge are perceptually significant in the point set clustering. In general, an *inconsistent edge* is an edge of MST whose weight is significantly larger than the mean of nearby edge weights on both sides of the edge (e.g. bold edges in Fig. 4). It can be defined as follows:

$$edge_i = \begin{cases} inconsistent, & if\ w_i > I_l \cap w_i > I_r \\ consistent, & otherwise \end{cases}, \quad (2)$$

where I is a measure that can be defined on both left (I_l) and right (I_r) sides of $edge_i$:

$$I = max\{f \cdot mean_{weight},\ mean_{weight} + n \cdot sd_{weight}\}\ .$$

An edge is said to be inconsistent if its weight exceeds the mean weight ($mean_{weight}$) of its neighboring edges (within p depth) on both sides by n units of standard deviation (sd_{weight}) and further more if it is at least f times as large as both 2.0n weights (Zahn, 1971, p. 82). Given a normal distribution, there is less than 1% chance that an edge weight would exceed $mean_{weight}$ by 3 unit of sd_{weight}. Therefore, $n \geq 3$ is regarded to be statistically significant. In some cases where sd_{weight} approaches to zero, the factor $f > 1$ ensures that the inconsistent edge is still 'outstanding' compared with $mean_{weight}$. A detailed discussion on the parameterization issue refers to Zahn (1971).

However, inconsistent edge alone is not sufficient for detecting building patterns. We did experiments and found that no matter how to adjust and combine the parameters, most of the building patterns do not automatically show themselves up by simply cutting the inconsistent edges off the MST except for some rough clusters. The reason is mainly that the MST-based techniques, which only enforce the proximity principle (Zahn, 1971), do not always lead to a result that resembles human perception. There-

fore other principles of perceptual organization must also be integrated for further processing depending on the kind of pattern to be detected. For the detection of collinear or curvilinear patterns, the principle of good continuity is much preferred.

The parameterization of inconsistent edge seems to be not very critical in our work. Since further processing is required for detecting specific patterns anyway, a conservative set of parameters was primarily used in this work (i.e. $p = 2, n = 3,$ and $f = 2.0$), making sure that some very significant edges are pruned while other less significant ones are kept for further decisions. However, our experiments show also that the use of other sets of parameter values with small variations makes little difference for the final detection results, as the difference caused in the pruning step can be made up for by refined processes (see Section 3.2).

3.2 Align-along-road building pattern
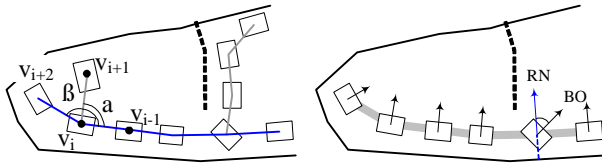
(a) concepts for the detection of align-along-road pattern   (b) concepts for the characterization of align-along-road pattern

Figure 5: Related concepts for detecting and characterizing align-along-road pattern

There is a general Graph-theoretic based framework for the detection of all linear alignments in our work, that is, the detection is achieved by tracing paths in the previously obtained MST, and the trace of the paths has to conform to different sets of constraints appropriate for the detection of each type of patterns. Formally, a *path* is defined here as a sequence of vertices in a MST (e.g. the bottom line going through the buildings in Figure 5(a)); and the *path angle* at vertex v_i is defined by the angle of $vector(v_{i-1}, v_i)$ and $vector(v_i, v_{i+1})$ in that path (Figure 5(a)).

Detection of the align-along-road pattern The basic idea of detecting align-along-road patterns is that a path should be traced from the pruned MST and the buildings on the path should be close enough to a nearby road. Therefore, we formulate the detection of the align-along-road pattern as a path tracing procedure that advocates the same aligned road, proximity, size similarity and path angle constraints (Algorithm 1). This choice of constraints set is justified as follows.

As a result of studying existing maps, we find that shape and orientation are not as dominant as aligned road, proximity, and size in determining a collinear group. First, that buildings align along the same road can be recognized as a align-along road pattern is a fundamental constraint, which should never be violated. Second, most buildings have similar man-made shapes; those with complex shapes are usually much bigger than the ordinary buildings, and thus they can easily be filtered out by size similarity in the first place. In addition, a generic shape measurement for map generalization is still not available; commonly used shape measures (e.g. compactness, shape index) describe specific aspects of shape and thus are not sufficient for our case. Third, the theoretically maximum deviation of building orientation computed with the wall statistical weighting method is $45°$ (Duchêne et al., 2003), which has much less impact on the perception of align-along-road patterns than proximity constraint. Consequently, we discarded the use of shape and orientation constraints in our detection algorithm of align-along-road pattern in the experiments. Nevertheless, these two aspects are integrated in the characterization of this pattern.

Algorithm 1: Detecting the align-along-road pattern

Input: buildings; partition roads; $constraints$
Output: $Collection$ of align-along-road patterns ($AARP$)

1: calculate CDT and then MST for the input data;
2: prune the inconsistent edges from the MST;
3: trace paths in the pruned MST as follows:
foreach v_0 = *vertex of degree 1 or vertex of degree 3 in MST* **do**
  loop "initialize $AARP$"
    v_1 = adjacent vertex of v_0;
    if v_0, v_1 *align along the same closer road* **then**
      initialize $AARP$ with $edge(v_0, v_1)$;
      break;
    else
      $v_0 = v_1$;
  endloop;
  loop "add new buildings to $AARP$"
    if v_1 *is of degree 1* **then**
      add $AARP$ to $Collection$ if $AARP$ contains more than 3 buildings;
      break;
    if v_1 *is of degree 3* **then**
      select a vertex v_2 with which the $edge(v_1, v_2)$ forms larger path angle with its predecessor edge.
    else
      v_2 = adjacent vertex of v_1
    check $edge(v_1, v_2)$ w.r.t. the $constraints$ for $AARP$;
    if $edge(v_1, v_2)$ *exists* **then**
      add $edge(v_1, v_2)$ to $AARP$;
      $v_0 = v_1$; $v_1 = v_2$; v_2 = NULL;
    else
      add $AARP$ to $collection$ if $AARP$ contains more than 3 buildings;
      empty $AARP$; renew $AARP$ with $edge(v_1, v_2)$ if the two vertices share a road;
      $v_0 = v_1$; $v_1 = v_2$; v_2 = NULL;
  endloop;
4: combine the detected $AARP$ if possible;

The mechanism how the selected constraints work is described as follows. These constraints work on a new edge occurring in the tracing process in 1, determining if the new edge can be added to an existing alignment. Aligned road is firstly checked using the information stored as a result of constructing refined CDT. If buildings connected by the new edge share the same road, then the tracing proceeds; and vice versa. The proximity constraint is based on the idea of inconsistent edge introduced in Section 3.1: if the weight of the new edge is inconsistent concerning the existing pattern, it cannot be added to the existing pattern. Finally, the size similarity constraint ensures that the size contrast between the two buildings at both ends of the new edge should not be too large (i.e. bigger building/smaller building $< size\_contrast$). It has to be mentioned that in the iteration step where there are more than one new edges available (e.g. $edge(v_i; v_{i+1})$ and $edge(v_i; v_{i+2})$ in Figure 5(a)), an edge with larger path angle ($\alpha, \beta \in [0°, 180°]; \beta > \alpha$) should be selected and added to the existing pattern in order to keep good continuity principle (Wertheimer, 1923).

In the step 4 of Algorithm 1, a combination is recommended in the cases of two detected groups approach to each other at their ends. This combination can be done by introducing a connecting edge consisting of two vertices at the proximate ends of both groups, if on the one hand the two groups align along the same road and on the other the distance of the introduced connecting edge is not too long. The parameter showing promising results

Algorithm 2: Detecting unstructured clusters	

Input: buildings; partition roads [optional]; $postconditions$
Output: $Collection$ of unstructured clusters (UC)

1: calculate CDT and then MST for the input data;
2: prune the inconsistent edges from the MST;
3: prune the edges connecting buildings whose size difference $> size\_contrast$; if $FragRatio > 0.5$, the edges should not be pruned;
4: organize connected subgraphs and populate UC with them;
5: filter out those UC that cannot pass $postconditions$, and add remaining UC to $Collection$;

for align-along-road pattern (Section 4) is $size\_contrast = 3.2$.

Characterization of the align-along-road pattern The characterization of align-along-road building patterns is by applying Equation 1 to spacing, size, shape, and distance to the aligned road. For these homogeneities, mean values are firstly calculated for the properties and then the homogeneity is computed. It is worth noting that both spacing (i.e. inter distance between buildings) and distance to the aligned road are computed using the nearest distance calculation forth mentioned in Section 3.1.

The calculation of the homogeneity concerning the orientation is described as follows. The orientation is considered to be more regular if the buildings change their orientations right according to the normal directions (e.g. RN in Figure 5(b)) of the local road segments that they align. For each building in the pattern, an angle deviation $\in [0°, 45°]$ is calculated between RN (Normal of Road segments) and BO (Building Orientation), the calculation of the two is presented in Section 3.1; the homogeneity of orientation is then computed from the standard deviation of all the deviations.

3.3 Unstructured clusters

Detection of unstructured clusters Unstructured clusters are also a common feature on topographic maps, especially at larger scales (1:10k-1:50k).

The method (Algorithm 2) detecting this type of building pattern is realized simply through pruning edges that are inconsistent and edges that connect two buildings whose size difference exceeds $size\_contrast$, and grouping the connected subgraphs from the pruned MST. Those subgraphs, however, are just candidates which have to tested against several $postconditions$. t The first $postcondition$ is the number of buildings contained. in our experiments, we define that at least three buildings form a cluster or pattern, as only in this case the calculation of mean and standard deviation required by the detection and characterization is meaningful. A second $postcondition$ is so-called black-and-white ratio, which is defined as follows:

$$BWRatio = \frac{\sum Area(b_i)}{Area(ConvexHull(UC))} \ ,$$

where $b_i \in UC$. This $postcondition$ protects some wriggling linear alignments from being recognized as unstructured clusters (Figure 6(a)). A final $postcondition$ is termed fragmentation ratio ($FragRatio$), which is the ratio between the number of pruned edges and the number of total edges in an initial cluster. This $postcondition$ should be applied in step 3 (see Algorithms 2) to prevent clusters from being over-fragmented, as in the cases where small and big buildings are arranged alternately (e.g. Figure 6(b)).

Characterization of unstructured clusters Despite the homogeneity of orientation which is calculated based on absolute ori-
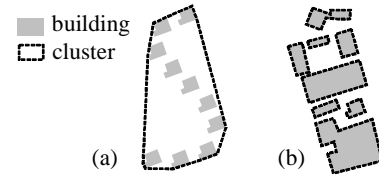

Figure 6: Bad examples of unstructured clusters

entations of buildings, the homogeneity of spacing, size, shape is exactly the same to those calculated for align-along-road patterns. The algorithms for the detection and characterization were implemented and results are presented in Section 4.
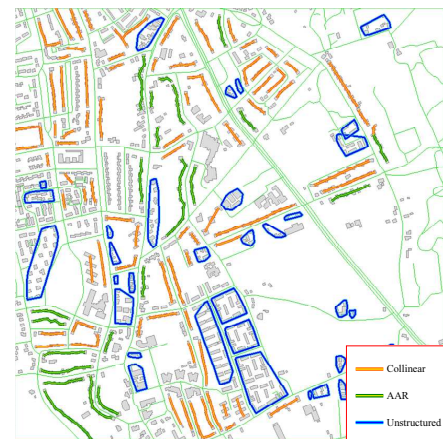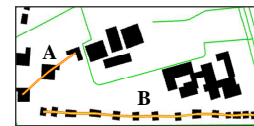
4 IMPLEMENTATION AND RESULTS

Figure 7: Test case and detection results of collinear, AAR, and unstructured patterns

id	spacing	size	orientation	shape
A	0.06	0.32	9.74°	0.43
B	0.52	0.26	5.57°	0.11

Figure 8: Measured characteristics of collinear alignments

We implemented the two proposed algorithms in an interactive generalization prototype system using C++. In addition, an algorithm detecting collinear building patterns was also implemented. The test case and detection results for collinear, align-along-road (AAR), and unstructured patterns are shown in Figure 7. There are 151 patterns detected from the test dataset, where collinear alignments are 94, unstructured clusters are 34, and align-along-road patterns are 23. It is noticeable that some building groups can be e.g. both AAR and collinear alignments. In this implementation, the final pattern type of a cluster was decided according to the computed characteristics; the pattern type with smallest homogeneity values was assigned to the cluster.

The characterization results are shown in Figures 8, 9, and 10. A general observation is that the measure values for spacing, size, shape, and dis2road (values $\in [0, 1]$) and the values for orientation (values $\in [0°, 45°]$) are consistent with our perception of these detected patterns. For examples, the measured homogeneities of size, orientation, and shape for the collinear patterns in Figure 8 shows that alignment B is superior to A in these aspects; the homogeneous cluster shown in Figure 9 is confirmed
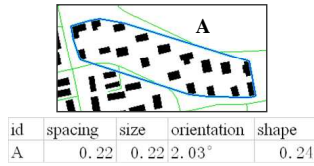
id	spacing	size	orientation	shape
A	0.22	0.22	2.03°	0.24

Figure 9: Measured characteristics of an unstructured cluster

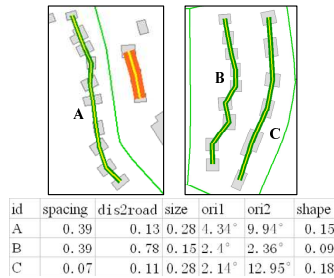| id	spacing	dis2road	size	ori1	ori2	shape
A	0.39	0.13	0.28	4.34°	9.94°	0.15
B	0.39	0.78	0.15	2.4°	2.36°	0.09
C	0.07	0.11	0.28	2.14°	12.95°	0.18

Figure 10: Measured characteristics of align-along-road patterns

by the measured homogeneities of this unstructured cluster, as all the homogeneity values are relatively low (below a quarter).

For the align-along-road patterns as shown in Figure 10, several observations can be made. First, 'dis2road' column confirms that this characteristic is more homogeneous for A and C than B, as B is aligned along the road to its left. Second, an extra orientation ('ori2') is calculated based on absolute values, in order for the readers to compare it with the orientation change with aligned roads ('ori1'). The results show that 'ori1' is generally less than 'ori2', as we can also observe in Figure 10 that although the individual orientations are rather fluctuated the orientations indeed change according to their aligned roads respectively. This observation confirms that orientation varying according to the aligned road is well suited for characterizing align-along-road pattern.

5 DISCUSSION AND CONCLUSION

This paper proposes a typology of building patterns available in cartographic and geospatial domain based on the study of existing topographic maps and related literatures. In this typology, fundamental visual patterns have been identified and their characteristics been discussed. This is an important step towards formalizing our knowledge on the building patterns in this field. Then, two Graph-theoretic based algorithms are presented in order to detect two pattern types of the typology, namely the align-along-road and unstructured patterns. The detection and characterization method was implemented and tested using Dutch topographic datasets. The results appear to be promising.

The proposed approach is generic in the sense that visually important building patterns can be detected and characterized no matter the buildings are topologically adjacent or not, even if the spatial objects are represented by points, since all these cases can be handled by the refined CDT. This approach can be further applied to detecting the patterns of archipelago, by adjusting some of the constraints or postconditions of the algorithms.

Although we argue in this paper that in most cases collinear and curvilinear alignments can be replaced by align-alogn-road patterns, the characterization of align-along-road patterns is still too general. That is, some important characteristics, like the straightness and main angle of collinear alignments, and smoothness and curvature descriptions of curvilinear patterns, would be lost if those linear patterns were recognized as align-along-road patterns. Therefore, future work will be focus on the detection and characterization of curvilinear alignments.

Further testing of the proposed approach against non-Dutch datasets will be carried out in order to confirm the claimed generality. Also noted that another work on the automated evaluation of building pattern preservation constraint (Zhang et al., 2010) is carrying out based on the detection results reported in this paper.

REFERENCES

AGENT, 1999. Selection of basic measures, deliverable c1. Technical report.

Anders, K.-H., 2006. Grid typification. In: Progress in Spatial Data Handling, pp. 633–642.

Boffet, A. and Rocca Serra, S., 2001. Identification of spatial structures within urban block for town qualification. In: ICC, Vol. 3, Beijing, pp. 1974–1983.

Christophe, S. and Ruas, A., 2002. Detecting building alignments for generalisation purposes. In: D. E. Richardson and P. van Oosterom (eds), Advances in Spatial Data Handling (SDH 2002), Springer Verlag, Berlin, pp. 419–432.

Duchêne, C., Bard, S., Barillot, X., Ruas, A., Trévisan, J. and Holzapfel, F., 2003. Quantitative and qualitative description of building orientation. In: 5th Workshop on Progress in Automated Map Generalization, Pairs.

Lüscher, P., Weibel, R. and Burghardt, D., 2009. Integrating ontological modelling and bayesian inference for pattern classification in topographic vector data. Computers, Environment and Urban Systems 33(5), pp. 363–374.

Regnauld, N., 1996. Recognition of building clusters for generalization. In: M. J. Kraak and M. Molenaar (eds), Advances in GIS Research II (Proceedings of 6th SDH'96, Delft), Taylor & Francis, London, pp. 4B.1–4B.14.

Ruas, A. and Holzapfel, F., 2003. Automatic characterisation of building alignments by means of expert knowledge. In: ICC, Durban, pp. 1604–1515.

Stoter, J., Burghardt, D., Duchene, C., Baella, B., Bakker, N., Blok, C., Pla, M., Regnauld, N., Touya, G. and Schmid, S., 2009a. Methodology for evaluating automated map generalization in commercial software. Computers, Environment and Urban Systems 33(5), pp. 311–324.

Stoter, J., van Smaalen, J., Bakker, N. and Hardy, P., 2009b. Specifying map requirements for automated generalization of topographic data. Cartographic Journal, The 46(3), pp. 214–227.

Wertheimer, M., 1923. Laws of organization in perceptual forms. In: W. D. Ellis (ed.), A Source Book of Gestalt Psychology, Routledge & Kegan Paul, pp. 71–88.

Zahn, C. T., 1971. Graph-theoretical methods for detecting and describing gestalt clusters. Computers, IEEE Transactions on C-20(1), pp. 68–86.

Zhang, X., Stoter, J., Ai, T. and Kraak, M.-J., 2010. Formalization and data enrichment for automated evaluation of building pattern preservation. In: Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science (SDH' 2010). to appear.