Extraction of Building Polygons from SAR Images: Grouping and Decision-Level in the GESTALT System

Eckart Michaelsen[1], Uwe Stilla[2], Uwe Soergel[3], Leo Doktorski[1]
[1]*FGAN-FOM, Gutleuthausstrasse 1, Ettlingen, Germany*
[2]*Photogrammetry, TU Munich, Arcisstrasse 21, Munich, Germany*
[3]*Phottgrammetry, Leibniz Uni. Hannover, Nienburger Str. 1, Hannover, Germany*
mich@fom.fgan.de

Abstract

The GESTALT-System is a stratified architecture for challenging computer vision tasks. This contribution focuses on the 3rd and 4th layer of it – the grouping and decision layers. As example application building recognition from high resolution SAR-Data is presented. The 3rd layer contains an assessment driven perceptual grouping process with any-time capability and flexible control. Important grouping principles such as good continuation and symmetry are utilized. A dynamic programming optimization is used in the final decision and post-processing layer to find closed polygons that describe the outlines of buildings. Further post processing includes polygon editing and consistency enforcement.

1. Introduction

Modern SAR-sensors provide spatial resolutions at least one order of magnitude better than only one decade ago. Airborne systems can achieve geometric resolutions in the decimeter scale. Figure 1 shows an example picture. New opportunities such as looking for buildings in urban environments are given. In this contributuion a complex structural methodology - the GESTALT-system (Grouping Evidence System for Treatment of Alternatives within a Layered Task-solver) - is utilized. It emphasizes free combinatorial yet robust search for perceptual Gestalts in one of its intermediate layers, while postponing decisions as far as can be afforded.

Automatic “image understanding” has been a major issue in the pattern recognition and computer vision community for some decades [3]. A much elaborated example for rule-based reasoning systems has been the well known SIGMA-system [10]. The SCHEMA-system [6] was a well known example with still lasting impact on the computer vision community.

A comprehensive overview on reliable building extraction can be found in the Ascona workshops [8]. Urban regions were regarded as being particularly difficult. A very interesting and recent example is [12] where buildings are extracted from colored aerial imagery. For these data color and texture attributes of 2d segments can be utilized. Building recognition from SAR-data is considered e.g. in [1, 2, 9, 13]. Own previous work can be found in [11].

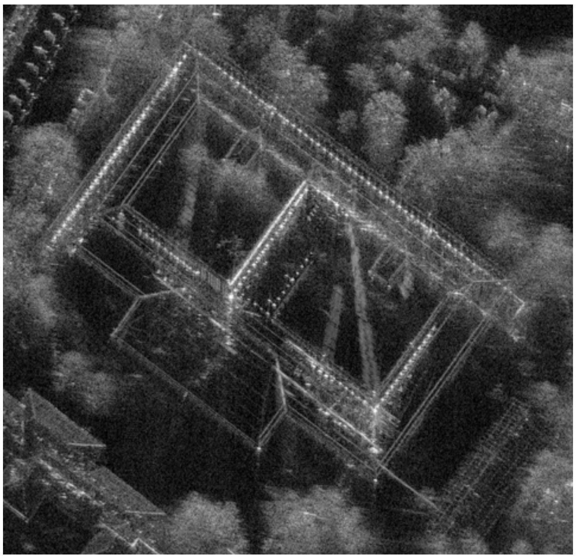

Fig. 1. Example of a high resolution SAR-image

Building recognition on standard SAR data was in general restricted to very big buildings. Most evidence was drawn from backscatter caused by roof structures or from the strong shadows appearing opposite to the illumination direction. In urban areas these features often cannot be applied, because of preferred smooth building facades and shortened shadow areas due to mutual signal interference with other buildings. In very high resolution SAR-images the appearance is dominated by salient edge and point scatterers. They are usually quite dark almost anywhere else. Fig. 1 shows an example of such an image. Some of the presented modules and methods – such as the use of morphological filtering in the preprocessing layer – may profit from particular SAR phenomena (only bright objects are of interest) but most of them are burdened from the specific SAR-phenomena – such as the high dependence of the appearance on the illumination direction and strong non-Gaussian noise. Most of the methods that are combined in the GESTALT system (Sects. 2 to 4) rely on properties not of the sensor but on general object features such as symmetry and repetitive structure.

2. The Overall Architecture of the GESTALT System

GESTALT has the following four layers:

Preprocessing: The data remain in iconic form. For the presented building recognition application two branches are present. In one branch the image is scaled down by factor 5 and than a morphological opening operator is applied to enhance and isolate spot structures. In the other branch the image is tiled into 25 over-lapping sub-images and on each of these a morphological closing is applied in order to close gaps and enhance thin line structures [14].

Feature extraction: Here the data are transformed into symbolic descriptions, i.e. a set of primitive objects. These get positions, orientations etc. attached as attributes. From the reduced image pixels that are brighter than the surrounding are obtained by a spot filter [own citation]. In the example there are 4275 such primitive components of type P.

From the 25 sub-images short line segments are extracted using the squared averaged gradient filter [7] and a threshold. Preliminarily, these are located at individual pixels. A first grouping according to co-linearity is performed in each sub-image separately. For the example the resulting set of primitive objects of type L contains 18842 elements from all 25 sub images.

Grouping: The mid-level grouping layer is described in more detail in Chapter 3. The mid level avoids decisions. Consistence enforcement is delayed to the next level of the system.

Decision and post-processing: This layer of the system is meant to produce the desired output. This layer is described in more detail in Chapter 4.

3. The Grouping Layer

In this example application the system operates on a DB consisting of objects of the following types: P, L, LL, Sc, A, R, Sy and C – i.e. pixels, short lines, long lines, scatterers, angles, rows of scatterers, symmetries and symmetry clusters. For these objects the productions are defined as given in Table 1.

Tab. 1. Production rules for grouping

LL	*regression*	*collinearity*	L … L
A	*intersection*	*proximity*	LL,LL
Sy	*axis & location*	*symmetry*	A,A
C	*centre*	*proximity*	Sy…Sy
Sc	*centre*	*proximity*	P…P
R	*initialization*	*proximity*	Sc,LL
R	*append*	*continuation*	Sc,R

While the grouping layer is active evidence is accumulated. There are no decisions and there is no enforcement of consistency. For each newly constructed object a hypothesis is formed. Preliminarily a *nil* is assigned to it. When such a *nil*-hypothesis triggers search it will be cloned using the knowledge given in Table 1 in the rightmost column. E.g. an object A will get the hypothesis assigned to be part of an object Sy. A non-*nil* hypothesis will trigger a query to the DB using the constraints given in the third column of Table 1. This will lead to a set of admissible partner objects. With these the construction (given in the second column) will be performed – resulting in new objects which again get *nil*-hypotheses attached. The administration of the set of current hypotheses is the core of the grouping process. Details are given in [11]. It is assessment driven. Sorting is not done after every hypothesis processing. Instead N cycles are processed en block before the next reorganization of the queue (e.g. N=100). Processing can also be performed parallel. For the experiments documented here 100.000 hypotheses were performed. Table 2 shows the numbers of objects reached after that..

Tab. 2: No of Objects

P	4275
L	18842
Sc	218
LL	8957
R	7511
A	13669
Sy	30845

4. The High-level Layer

The task of building recognition demands a specific result format. Buildings are outlined by closed polygons in GIS-systems. And there are constraints

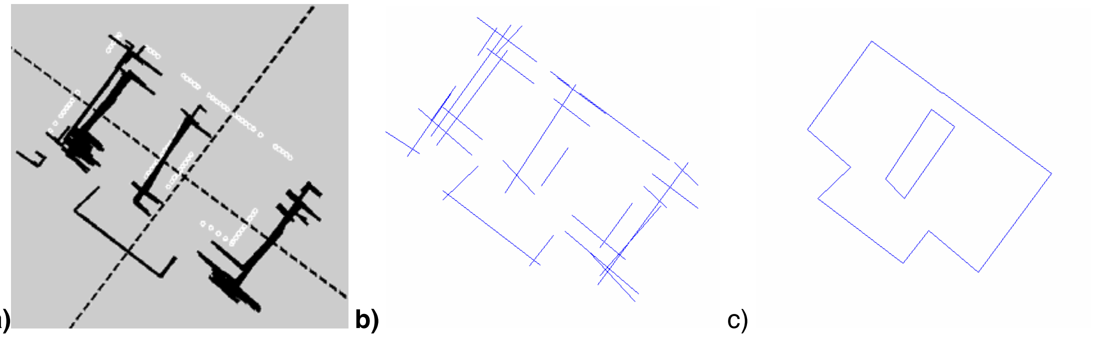

Fig. 2: a)best grouping objects, b)selected contours c) final poygons

demanding that building polygons cannot intersect each other or overlap. The last layer of the system transforms the objects resulting from the grouping layer into such format while enforcing the constraints. This is achieved by the following sub-steps.

1. Filtering the accumulated set: Only components *C* and *R* are of further interest. Components *C* are evaluated according to the number of components *Sy* in them and the consistency of the symmetry axis with the two preference orientations. For each orientation the one with the best evaluation is kept and all others are removed. For the example these two components C are displayed in black in Fig. 4) on the left side with the axis shown as dashed line and the preceding components *LL* as solid lines. These are regarded as contour objects. Objects *R* are evaluated according to the number of scatterers in them. Only those with more than three members are kept. Furthermore, all such objects that have a successor with more members are re-moved as well. The remaining rows are regarded as contours as well. They are displayed in Fig. 2a in white color. Among the set of contours again another grouping is performed. Basically, this uses the same co-linearity production that also applied in the grouping layer and even in the feature extraction layer on each sub-image. But decisions are not avoided anymore. In the example only 32 contours survive this filtering step. They are shown Fig. 2b.

2. Searching for polygon candidates: Given a set of contour line objects $S=\{s_1,\ldots,s_n\}$ the cyclic polygons are k-tuples taken from S with no repetition (where $k\leq n$) and an equivalence ~ allowing cyclic permutation and reversion. Each contour object s_i has two end points $x_{i,0}$ and $x_{i,1}$ and indices p, q are understood modulo 2 such that $p+1=0$ iff $p=1$. A function on pairs of consecutive contour objects in a polygon and $p=0,1$ is given as

$$(\mathrm{i},\mathrm{j},\mathrm{p}) \mapsto \left(C_p\left(s_i,s_j\right),q\right),$$

where the cost function is defined as

$$C_p\left(s_i,s_j\right)=\min_{q\in\{0,1\}}\left(\left\|x_{i,p}-x_{j,q}\right\|/\left\|x_{i,p+1}-x_{j,q+1}\right\|\right)$$

and $q=\Pi(p)$ is the argument of the minimization. Such functions punish large distances between end points $x_{i,p}$ and $x_{i+1,q}$ (close to each other) and rewards large distances between the other end points $x_{i,p+1}$ and $x_{i+1,q+1}$ (far from each other). The cost for a polygon is than given as sum

$$\mathbb{S}\left(s_{i_1},\ldots,s_{i_k}\right)=\sum_{v=1}^{k}C_{p_v}\left(s_{i_v},s_{i_{v+1}}\right),$$

where $p_v=\Pi(q_{v-1})+1$; i.e. if in a particular step the minimum was found for $q=0$ then in the next step $p=1$ and vice versa. Of course v is understood modulo k here. Such cost function can be minimized by dynamic programming approach with an additional binary matrix that records the choices q_v for each step and each contour object. Dividing the cost by the number of contour objects k gives the mean cost for a polygon. It is not guaranteed that the global minimum will be found for two reasons: 1) the starting index i_1 may actually refer to a clutter contour object that does not participate in any sensible polygon; 2) there is a greedy decision in the minimum at the definition of C_p. To alleviate this the search is repeated with $i_v=1,\ldots,n$ and start value $p_1=0,1$.

From the maximally $2n$ resulting optimal polygon candidates many can be equivalent with respect to ~. Only one representative of each of these equivalence classes must be kept and is subject to the following editing and decision procedures.

3. Editing polygons: Let $(s_1,\ldots,s_k)$ denote a candidate found by the dynamic programming search. The building recognition tasks demands closed polygons where the end point of each segment is the starting point of the next. These new vertices are calculated from each pair of successive line contour objects (s_l,s_{l+1}). There are three cases: 1) there is a significant difference in the orientation; then the corresponding vertex is calculated as intersection of lines; 2) (s_l,s_{l+1}) are almost collinear; then the vertex is found by averaging the closer ends; 3) (s_l, s_{l+1}) are almost parallel and not collinear; then a new segment is inserted between s_l and s_{l+1}. I.e. closeness is enforced by construction of new not observed virtual segments.

4. Enforcing constraints: Some of the edited polygons may not be contradictive: If they are completely detached they will indicate separate independent building entities. If they are completely inside each other the inner one may be a yard inside the outer one. However, some of them may be mutually contradictive. If two polygons intersect each other a decision will be needed, because two buildings cannot cover partially the same space. In some GIS formats buildings are allowed to touch each other, i.e. have common segments. We cannot tolerate this for the automatically constructed results at the present state of our research. Instead among a set of polygons that have common segments the one that covers the biggest area is chosen. Fig. 2c displays the two remaining polygons after these decisions. The outer big one roughly fits the true outline of the building. The smaller inner one fits to a true sub-structure on top of the roof of the building. The two inner yards are not instantiated.

5. Discussion and Conclusion

A very diverse set of image processing and recognition modules is organized in a systematical way so that their overall cooperation and competition leads to a preliminarily acceptable result on a difficult task. Early decisions are avoided and in the intermediate grouping layer the system is allowed to freely group many things together and even produce illusions. Success is owed much to sorting according to well balanced assessments. This concentrates the given computational resources on relevant work. Only in the grouping layer 3 with its intelligent control combinatorial complexity can be accepted. This can be run as long as is affordable. It has any-time capability. All other layers contain methods of low polynomial complexity at most. Their computational needs can be estimated in advance.

On the other hand the result is not yet really satisfactory – e.g. the two salient inner yards are missing. But, no knowledge about the specific nature of SAR images has yet been used – unlike almost all other work on buildings in SAR data such as [1, 2, 9, 12]. That is a very promising topic for future work.

Another important issue is top-down evidence: For example if in the decision layer high cost arise at a particular contour segment in order to enforce closeness it may be a good idea to query the mid-level DB again for the presence of components at the corresponding position or to start the feature extraction – differently parameterized – again in such a focus of interest area. We also expect major improvements from such feedback.

References

1. Balz, T., Haala, N.: Interpretation of High Resolution SAR Data using Existing GIS Data in Urban Areas. IAPRS Vol. 36 Part 3/W24 Joint Workshop of ISPRS and DAGM CMRT05 (2005) 111-116.
2. Bolter, R., Leberl, F.: Phenomenology-Based and Interferometry-Guided Building Reconstruction from Multiple SAR Images. In: Proceedings of EUSAR 2000 (2000), 687-690.
3. Crevier, D. Lepage, R.: Knowledge Based Image Understanding Systems: A Survey, CVIU, Vol. 67 (1997) 161-185
4. Desnos, Y. L., Matteini, V.: Review on Structure Detection and Speckle Filtering on ERS-1 Images. In: EARSeL Advances in Remote Sensing, Vol. 2 (1993) 52-65
5. Desolneux, A., Moisan, L., Morel, J.-M.: Gestalt Theory and Computer Vision. In: Carsetti, A. (ed.) Seeing, Thinking and Knowing. Kluwer, Dordrecht (2004) 71-101.
6. Draper, B., Collins, R., Brolio, J., Hanson, A., Riseman, E.: The Schema System, IJCV, Vol. 2 (1989) 209-250
7. Foerstner, W.: A Framework for Low Level Feature Extraction. In: Eklundh, J.-O. (ed). Computer Vision – ECCV 94. Vol. II, B1 (1994) 383-394
8. Gruen, A., et al. (eds.) Automatic Extraction of Man-Made Objects from Aerial and Space Images (I, II and III). Three workshops in Ascona; Birkhäuser, Basel, (1995, 1997 and 2001)
9. Gamba, P., Houshmand, B., and Saccini, M.: Detection and Extraction of Buildings from Interferometric SAR Data. IEEE Transactions on Geoscience and Remote Sensing. Vol. 38, No.1 (2000), pp. 611-618
10. Matsuyama, T., Hwang, V. S.-S.: Sigma a Knowledge-based Image Understanding System, Plenum Press, New York (1990)
11. Michaelsen, E., Soergel, U., Thoennessen, U.: Perceptual Grouping for automatic Detection of man-made Structures in High-resolution SAR Data. Pattern Recognition Letters, Vol. 27 (2006) 218-225.
12. Schuster, H.-F.: Detection of Man-made-objects Based on Spatial Aggregations, http://www.ipb.uni-bonn.de/ipb/ projects/OntoSkalen/html/TB-ipb-05_1.pdf (2005)
13 Simonetto, E., Oriot, H., Garello, R.: Radargrammetric Processing for 3-D Building Extraction from High-Resolution Airborne SAR Data, IGARSS 2003, IEEE, Toulouse (2003)
14. Soille, P., Pesaresi, M.: Advances in Mathematical Morphology Applied to Geoscience and Remote Sensing, IEEE Trans. on Geoscience and Remote Sensing, Vol. 40, No. 9 (2002) 2042-2055