

TOWARDS UNDERSTANDING URBAN PATTERNS AND STRUCTURES

E. Michaelsen, R. Gabler, N. Scherer-Negenborn

Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB,
Gutleuthausstrasse 1, 76275 Ettlingen, Germany
eckart.michaelsen@iosb.fraunhofer.de

KEY WORDS: Gestalt algebra, perceptual grouping, remotely sensed images, intelligent urban design

ABSTRACT:

Intelligent urban design is a set of principles for desirable future urban structures. Existing urban structures can be analysed using remotely sensed images. In order to foster this analysis both in speed and objectivity automation is proposed in this work. Automatic Gestalt perception is distinguished from automatic knowledge-based analysis. Both will be required. For the Gestalt side an algebraic approach is utilized. This Gestalt algebra operates on a 6-D domain containing position, orientation, frequency, scale and assessment. It defines how to form aggregates from parts. Any Gestalt can be combined with arbitrary others, but good assessments are only achieved, if the parts are mutually in Gestalt-arrangements. There are operations for mirror-symmetry, good continuation in rows and rotational-symmetry. In this paper experiments are made only with mirror-symmetry and row-continuation. Example images of Thimphu, Bhutan and Phoenix, Arizona are obtained by use of Google Earth. The results are to a large degree in accordance with human perceptual grouping. Some illusory groupings not in accordance with human perception, as well as examples salient to humans which are not instantiated by the system, are discussed as well.

1. INTRODUCTION

1.1 Urban Structures

(Benninger, 2001) lists ten principles of intelligent urban design: balance with nature, balance with tradition, appropriate technology, conviviality, efficiency, human scale, opportunity matrix, regional integration, balanced movement, and institutional integrity. Each is more specifically elaborated by examples on many scales from individual houses to provinces – yielding the pattern of an ideal, pleasant urban environment, more a goal than the description of an existing city anywhere in the world.

Today, remotely sensed images provide a huge mass of data of any city in the world. Such images are available on any scale between below meter per pixel resolution and kilometres per pixel resolution. In particular virtual globe systems, such as Google Earth, give a comfortable interface and a very rich database. So in principle it should be possible to identify deviations between such ideal and reality, and e.g. pick out specifically good and bad examples, in order to learn from them. Doing so straight-forwardly requires a lot of zooming in and out, looking, thinking, acquiring additional knowledge, and contemplation. Probably, some automation and tools would help. Automation will not only save time – it will also help in making the outcome less subjective.

1.2 Gestalt and Knowledge

Pre-attentive, unaware grouping of objects into probably relevant aggregates is referred to as Gestalt perception. Psychologists measured that e.g. mirror symmetry is a very fast and important grouping mechanism for pictorial data in human observers, see (Sassi et al., 2014; Treder, 2010).

In this paper we take the view that such grouping processes also contribute much to the understanding of urban patterns from remotely sensed images by human observers. Thus mathematical models and corresponding search procedures should be investigated, if automation of such understanding is the goal.

On the other hand, human understanding of urban patterns from remotely sensed images will also be guided by knowledge. Mostly, this guidance is less fast and can be accompanied by awareness, so that self-inspection or expert interviews may help in the construction of corresponding artificial intelligence systems. Mutual constraints and relations between meaningful objects in a plane can e.g. be captured by syntactic structures. Possible other machine-readable formats include semantic nets, ontologies, etc. Such formats allow automatic inference.

1.3 Related Work

Knowledge-based approaches to urban pattern understanding have a long history with a prominent example being (Matsuyama & Hwang, 1990). One other example of such early syntactic understanding systems working from pixel scale to nested urban structure of 1000 pixel scale is (Füger et al. 1992). In this millennium the interest is more in getting more robustness into such approaches by the use of learning and statistical inference (Zhu et al. 2009).

Some of our own previous work captured perceptual grouping according to Gestalt principles inside a knowledge-based approach – namely, in production systems (Michaelsen et al., 2010; Michaelsen et al., 2006). It is now one main intention of the contribution at hand to separate the perceptual – pre-attentive – grouping from the – artificial intelligence based – knowledge utilization.

2. GESTALT ALGEBRA

2.1 Domain

Following (Michaelsen, 2014) a Gestalt is a member of

$$G = \mathbb{R}^2 \times \mathbb{R} / \mathbb{Z} \times \mathbb{N} \times (0, \infty) \times [0, 1] \quad (1)$$

with the components of the domain named position, orientation, frequency, scale, and assessment. We are not in the knowledge-based remote sensing topics here, so we don't call our entities

‘building’ or ‘road’ or ‘roundabout’ or ‘city centre’. In avoidance of this the term ‘Gestalt’ is used. The meaning of position and scale attributes is self-evident. If a Gestalt g has frequency n it means to be self-similar with respect to rotations from S_n . The orientation then is the phase of such rotation with respect to some specified direction (e.g. East). Assessment = 0 means ‘meaningless’, while assessment = 1 means ‘almost surely meaningful’.

2.2 Operations

Three operations are given for the Gestalt domain in (Michaelsen & Yashina, 2014): A binary mirror operation

$$|: G \times G \rightarrow G : (g, h) \mapsto g|h \quad (2)$$

yielding assessment 1 for $g|h$ if g and h are mutually in mirror symmetry, proximity, and of similar scale, and both have also assessment 1. Otherwise the assessment of $g|h$ will be smaller than 1, but the operation is still defined. Aggregate Gestalten formed by such term all have frequency 2.

An n -ary row-forming operation

$$\Sigma : G^n \rightarrow G : (g_1 \cdots g_n) \mapsto \sum g_1 \cdots g_n \quad (3)$$

yielding assessment 1 for $\sum g_1 \cdots g_n$ if $g_1 \cdots g_n$ form a perfect row, are mutually in proximity and of similar scale, and all have also assessment 1. Again any deviations lead to smaller assessment of the aggregate. Aggregate Gestalten formed by such term all have frequency 2.

An n -ary operation preferring rotational patterns

$$\Pi : G^n \rightarrow G : (g_1 \cdots g_n) \mapsto \prod g_1 \cdots g_n \quad (4)$$

yielding assessment 1 for $\prod g_1 \cdots g_n$ if $g_1 \cdots g_n$ form a perfect rotational mandala, are mutually in proximity and of similar scale, and all have also assessment 1. Again any deviations lead to smaller assessment of the aggregate. Aggregate Gestalten formed by such term have frequency n .

For the details of the operations and proofs of algebraic closure we refer to (Michaelsen & Yashina, 2014). There is, however, one change: Formulae (4) and (9), or (4) and (10) of (Michaelsen, 2014), give the proximity-to-scale component of the assessment functions in a similar manner as (6) of this paper. Such function gives the same value for a ratio r as for $1/r$. This turned out being less in accordance with human proximity grouping. We replaced it by

$$as_p = \sqrt{e} \cdot r \cdot \exp\left(-\frac{r^2}{2}\right) \quad (5)$$

where r is again the ratio between position distances and mid-scale of the part Gestalten at hand. The root of the Euler constant e is added here in order to have maximal value $as_p=1$ for $r=1$. This function resembles in form the density of a Rayleigh distribution. The tail of it for $r>1$ has less mass than the original formulae of (Michaelsen & Yashina, 2014).

Unfortunately, Gestalt algebra for the time being gives no operation for 2-D-grid symmetries. Such operation might be appropriate for scenes like the Phoenix images below.

2.3 Search

Primitive Gestalten are extracted from the images by use of key-point detectors such as SIFT, or other appropriate methods such

as super-pixel segmentation. Search strategies then build higher order Gestalt-terms using the operations above recursively. Depending on the desired output, often a cluster-step gives the final output, because the best Gestalten are often found in multiple, slightly different variants.

3. SOME EXPERIMENTS

In context with the principles of urban design according to (Benninger, 2001) often the city of Thimphu in Bhutan is mentioned. For comparison a typical automobile-dominated city was chosen – Phoenix, Arizona.

3.1 Image Capture

The images were obtained from interesting structures there using Google Earth virtual globe system – see Table 1 for the Geo-coordinates and the following Figures for examples. The 3D-features of the virtual-globe system were deactivated, nadir view direction chosen, and the camera-to-ground distance was set mostly to 500m, giving a pixel size of 0.55m on the 1100×1040 images. For comparison also images of larger scale were included as can be seen from Table 1. Logos and Geo-coordinates etc. have not been removed – they should not make much difference.

3.2 Primitive Extraction

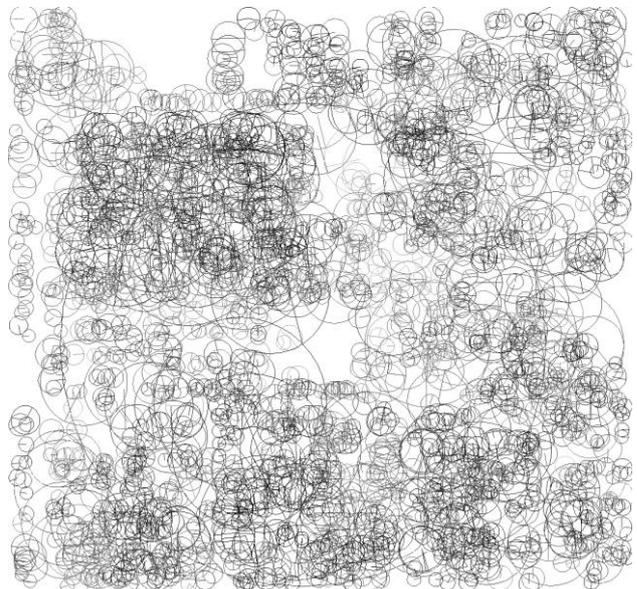


Figure 1. A set of SIFT primitive-Gestalten as obtained from Image #9, grey tones code assessment

Standard SIFT in MATLAB implementation was used yielding key-points and descriptors in numbers indicated in column 2 of Table 1. The key-points were used as primitive-Gestalten. Some example is displayed in Figure 1, showing the 12619 instances obtained from picture #9 (also used in Figure 3). The acceptance of a SIFT-key-point is based on a threshold on eigenvalue-properties of the outer product of the smoothed gradient. Thus an assessment can be given for every instance, giving value zero (indicated as white colour in the Figure) if the threshold was just met, and one for the maximal instance in that picture.

As the abbreviation SIFT already indicates, this extractor is strictly scale invariant – a property which should be softened for

#	N	E	Elev.	#SIFT	#(as>0.75)	#(as>0.9)	max(#row)	max(as)
Thimphu, Bhutan								
1	27°28'22.90''	89°38'15.65''	500m	23171	73	0	8	.8410
2	27°27'52.42''	89°38'28.88''	500m	21289	315	0	11	.8855
3	27°26'30.81''	89°39'59.21''	500m	21710	17	0	11	.8264
4	27°27'16.06''	89°39'17.74''	500m	22211	158	0	7	.8693
5	27°28'40.91''	89°37'22.24''	500m	22363	174	0	8	.8955
6	27°28'18.72''	89°37'51.12''	2000m	24277	146	0	9	.8285
Phoenix, Arizona								
7	33°28'14.32''	-112°02'03.80''	500m	18966	305	0	8	.8363
8	33°28'00.42''	-112°02'24.27''	500m	15373	233	0	8	.8722
9	33°27'16.73''	-112°01'38.61''	500m	12619	216	2	11	.9036
10	33°27'20.28''	-112°02'28.48''	500m	14193	254	0	8	.8577
11	33°37'01.86''	-112°17'23.56''	500m	15619	247	1	8	.9052
12	33°27'25.04''	-112°02'25.54''	2000m	24246	139	0	9	.8502
13	33°36'56.11''	-112°17'05.00''	2000m	27910	478	1	8	.9059
14	33°28'49.77''	-112°03'18.06''	8000m	19015	81	0	8	.8159

Table 1. Example pictures, Geo-reference, and resulting numbers and properties of Σ -Gestalten

remote sensing data. Recall that for objects such as buildings a rough scale is known. Therefore, a preference for primitives of a certain size sc_{pref} was introduced by re-assessing using:

$$as_{new} = as_{old} \cdot \exp\left(2 - \frac{sc}{sc_{pref}} - \frac{sc_{pref}}{sc}\right) \quad (6)$$

So the new assessment as_{new} equals the old as_{old} only if the scale of the primitive Gestalt sc equals the preferred value, and large deviations are punished. For the experiments reported here $sc_{pref}=50Pixel$ was set. In order not to overload the search only the 1200 best assessed Primitives are kept.

3.3 Search

The 128-dimensional SIFT-descriptors were also used to improve the assessments of the depth-one Gestalten, i.e. $|$ -Gestalten from primitives and Σ -Gestalten from primitives. This follows (Michaelsen, 2014). For the $|$ -Gestalten all pairs of input Gestalten are evaluated. For the Σ -Gestalten the same greedy search strategy was used as in (Michaelsen, 2014), which starts also from pairs, but then concentrates on the best continuation. Only rows with more than 2 members were kept. For this paper no rotary $||$ -Gestalten were investigated. The number of higher order Gestalten of each kind is limited to 500, again in order not to overload the current MATLAB-implementation. In this paper the search-depth was limited to 2. Figure 2c shows, that starting from primitives of our preferred scale sc_{pref} depth-2 Gestalten already fill large portions of the image.

3.4 Results

In the following list the results are discussed qualitatively. There is not enough room to present all results in a pictorial way, so this is done only for selected examples. Assessment statistics are given for all pictures in the tables 1, 2, and 3.

#1: centre of Thimphu, rather irregular; neither row operation nor mirror operation yield results in accordance with human perceptual grouping, Table 1 shows low assessments for the rows, this sets the clutter level, about .85 is appropriate; Table 2 gives almost the same for mirror grouping, however, here two of the best instances are located on major Buddhist shrines, and those are well in accordance with saliency to humans; Table 3 shows general higher assessments for level-2 Gestalten – in this

case mirror-of-row Gestalten; here the clutter level threshold should be set higher, e.g. to .95.

#	#(as>.75)	#(as>.9)	max(as)
Thimphu, Bhutan			
1	6	0	.8367
2	112	0	.8570
3	7	0	.8098
4	11	0	.8151
5	15	0	.8339
6	2	0	.7900
Phoenix, Arizona			
7	3	0	.7883
8	5	0	.8100
9	31	0	.8635
10	38	0	.8812
11	85	0	.8868
12	22	0	.8271
13	300	0	.8806
14	2	0	.7958

Table 2. Resulting numbers and properties of $|$ -Gestalten

#	#(as>.925)	#(as>.95)	max(as)
Thimphu, Bhutan			
1	35	0	.9372
2	14	0	.9368
3	0	0	.8995
4	78	1	.9508
5	76	4	.9535
6	75	0	.9358
Phoenix, Arizona			
7	207	0	.9459
8	88	0	.9434
9	175	4	.9543
10	332	0	.9477
11	203	1	.9628
12	25	0	.9410
13	500	112	.9620
14	35	0	.9318

Table 3. Resulting numbers and properties of $\Sigma||$ -Gestalten

#2: containing newly build residential house rows; most of the rows of similar new residential houses are found in accordance with human perceptual grouping, long $\sum P \dots P$ rows (up to eleven members) are instantiated on the salient saw-tooth roofed workshop, also many of the neat symmetries of the houses are instantiated as P/P in good accordance with human grouping, and with assessment higher than .85, some row-Gestalten above .85 are found along major straight contours along a road and a sports-field not in accordance with human perception, Table 3 shows no mirror-of-row Gestalt above .95 on this picture – the best one is actually on the overlaid logos and text;

#3: larger symmetric buildings – according to Google a “royal institute of managing”; one of the complexes is instantiated by a large cluster of P/P -Gestalten including the best one, but even this one has assessment far below .85, the other salient symmetric building is missed completely, row grouping fails to instantiate the salient triple row of buildings to the west of the institute, best rows are on the logos overlaid to the image, Table 3 shows also low assessments for the depth-2 Gestalten, all in all this example image yields almost complete failure of the method probably due to low contrast;

#4: again salient newly-built residential house rows; in accordance with human perceptions these are instantiated mostly, some rows are oblique and thus not in accordance with human grouping, some illusory row Gestalten are instantiated (e.g. on the logos or on a long straight contour at the bank of a river), and assessments are over clutter assessment but not very far; though the buildings are very symmetric automatic mirror grouping fails on this image, best $|$ -Gestalten are far below threshold and illusory (on the overlaid script), probably the lighting and shadow casting breaks the symmetries; the best level 2 symmetry of rows Gestalt is just above .95, but it is located as well on the overlaid script;

#5: very salient group of residential buildings; some results on this picture are given in Figure 2 displaying primitives in green, rows in red, and mirrors in blue; Table 1 already indicates that highly assessed $\sum P \dots P$ -Gestalten are instantiated here, and these are in high accordance with human perception; on the other hand P/P -Gestalten do not reach $>.85$ assessments, and in fact the individual buildings are not very symmetric; however, in this case, higher order grouping leads to a satisfying result, as can be seen in Figure 2c, where the only mirror-of-rows trespassing .95 threshold is displayed;

#6: overview over almost all of Thimphu; human perceptual grouping gives here only a large irregular cluster stretching along some major roads (or valleys), the tables indicate that nothing above clutter-level is grouped by Gestalt algebraic operations here, since all maximal assessments are far below clutter level this example can be regarded as a good ‘true negative’;

#7: residential area in Phoenix organized in rectangular blocks of about 198m×102m (where all of that city is clearly organized in square mile tiles – i.e. larger than the images at hand), the individual, probably single-family homes are placed in rows but mutually quite dissimilar in appearance, to human perception the street-grid is most salient, Gestalt algebra fails to instantiate anything like that with sufficient assessment, maximal assessments all fall far below the clutter thresholds, yet among the best assessed row- and mirror-Gestalten most are in some accordance with human vision;

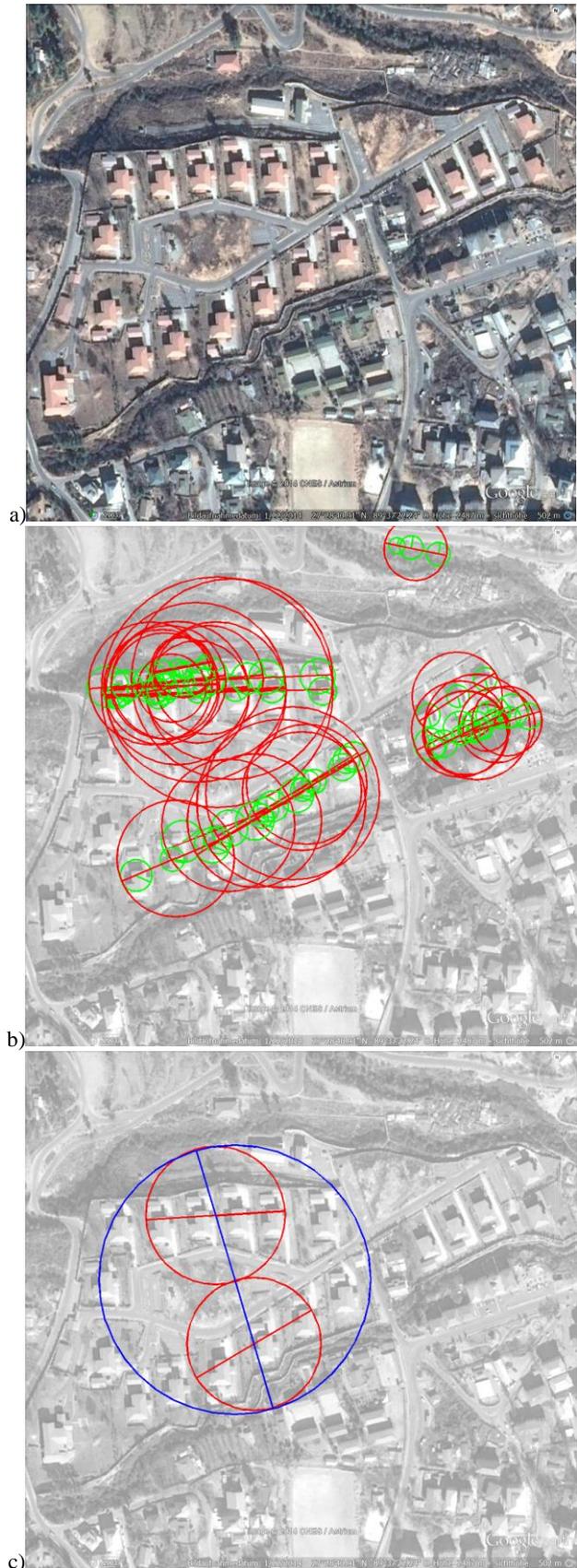


Figure 2. Thimphu scene #5: a) image as taken by virtual globe system; b) rows of primitive Gestalten $\sum P \dots P$ surpassing assessment 0.8; c) the best mirror-of-row Gestalt

#8: to the east and south this picture shows major motorways crossing, the rest is mostly residential area made up of single-family homes arranged in a square also salient to human vision, the system yields below clutter-level assessments for mirror- and second-level-Gestalten, Table 1 shows above level .85 Σ -Gestalten in this image, about half of those are in accordance with human grouping, the others are arranged along long contours given by the motorways, some are also obliquely oriented (where human perception will prefer either horizontal or vertical rows on this picture);

#9: this picture is displayed in Figure 3a, most salient to humans is a triple row of large administration buildings in the north western part, this one is also found by the Gestalt algebra search as maximally assessed Σ -Gestalt as can be seen on Figure 3b), threshold .80 was used in the level-1 Figures in order to show also false positives and what would still be missing even at this threshold, it can be seen that false rows appear e.g. on parking grounds and on salient contours, Figure 3c) exhibits that in picture #9 mirror grouping concentrates on the building complex in the south of the image which happens to be a prison, indeed this complex is also saliently symmetric to human perception;

#10: the large building complex in the south-west of this picture (a medical centre) has neither salient mirror-symmetry nor repetitions in rows, accordingly Gestalt grouping avoids it (although it has high contrast) and concentrates on the residential buildings in the rest of the image, either small row-houses or multiple-family homes, much of the automatic Gestalt grouping there is at least in partial accordance with human perception;

#11: residential area arranged in concentric circles around a shopping centre; this is a fairly new area obviously designed by repeating the same double-family homes along the curved streets, in accordance with human perceptions the Gestalt assessments are very high here – in fact top-level in all three Tables, since the rows bend along the curves Σ -grouping follows only about four instances long, then often two successive such level-1 Gestalten are combined into level-2 Π -Gestalten, which is the best the system can do in accordance with human perception, for the time being no Π -Gestalt grouping was tested, for this image this should be the proper operation;

#12: is an overview containing the hospital area of image #10 as well as the salient triple complex of #9, the latter is found again as top-level Σ -Gestalt (though the scale is four times larger here), some other high level rows are in accordance with human perception as well, while some are located along contours of a very big motorway crossing – not in accordance with human perception.

#13: same scale as #12 but containing the artificial rotational patterns of #11 which are very salient to human perception, in accordance to this the assessments found by the automatic Gestalt algebra system are also very high as can be seen from the tables, some of the top rated level-2 groupings are mirrors of rows repeating on a larger scale the same construction as in #11, but many are also illusory, almost all high rated level-1 mirrors are in accordance with human perception, but many are missing, highly assessed level-1 row Gestalten seem a little arbitrary;

#14: again a factor four in scale (8m per pixel), so that almost all Gestalt disappears, this is in accordance with the very low assessments found by the automatic system – as can be seen in the tables, the residual top-rated Gestalten either rest on the overlaid logos and script, or are illusory.



Figure 3. Phoenix scene #9: a) image as taken by virtual globe system; b) rows of primitive Gestalten $\Sigma P \dots P$ surpassing assessment 0.8; c) Π -Gestalten on the same image

4. DISCUSSION

The experiments show that many perceptual groups are found by the automatic search using Gestalt algebra in good accordance with human perception. Setting a suitable clutter threshold – as done above – will give next to those ‘true positives’ also some ‘false positives’, i.e. illusory groups, which are not in accordance with human perception. These are often on salient long straight contours, a fault more due to the key-point detector which obviously gives too highly assessed responses at such locations. So by improving the key-point method this should be avoidable. ‘False negatives’ i.e. Gestalten salient to human observers which are not instantiated above the clutter threshold by the automatic system have also been found. An example is given by image #3. This image happens to have low contrast, and some of the saliency to human observers results from colour-contrasts. Recall that colours are not used by the system up to now, which gives a possible future add-on that may well contribute in a way similar to the aid of the SIFT-descriptors.

It is also evident e.g. from Figure 2c) that deeper level structures can be automatically found in accordance with human perception, Gestalten of Gestalten. However, as yet, correctly such instantiated hierarchies remain shallow and not very frequent. For the time being, the threshold has to be set much higher for this – .95 as opposed to .85 for the Gestalten of primitives. This is partly due to the use of the SIFT-descriptors modifying the assessments. It is planned for future work also to incorporate such additional assessment components also into the deeper level Gestalten, next to other constraints. But then such nice results as in Figure 2c) might well be lost: recall that here a row of four members is set in mirror symmetry with a row of three members, and that the descriptors must be mirrored, which would lead to bad assessments on that instance – also because shadows will not obey the mirror rules.

Shadows are one example where knowledge about the imaging circumstances and the likely nature of objects in the scene may help, and thus the question arises how such knowledge can be used in cooperation or competition with the Gestalt algebra.

Gestalt grouping alone will anyway not give appropriate understanding so far as intelligent urban design is concerned. But the experiment shows that man-made design according to some plan or principle can to some degree be distinguished automatically from rather arbitrary urban structure. Whether the Gestalten then represent a beautiful Buddhist shrine or a symmetric American prison cannot be distinguished, also not whether it is a well-balanced several-families-home according to Benninger’s principles or a serious waste of space for a single family needing three cars to get along.

It is necessary to build some recognition on the obtained Gestalten, or to have the Gestalten instantiation working in cooperation with it. E.g. on a smaller scale some vehicle detector will help to identify certain Gestalten as vehicle rows, and thus assign the meaning ‘parking lot’. Those detectors may be trained by representative example material or model-based. And of course a road recognition method will help to measure the amount and importance of individual motor traffic present at the urban structure at hand.

ACKNOWLEDGEMENTS

We thank GoogleEarth and CNES/Astrium for providing the images.

REFERENCES

- Benninger, C., 2001. Principles of Intelligent Urbanism. *Ekistics*, 69(412), pp. 39–65.
- Füger H, Jurkiewicz K, Lütjen K, Stilla U. 1992. Ein wissensbasiertes System für die automatische Bildanalyse. In: Fritz LW, Lucas JR. (eds): *ISPRS, XVIIth Congress*. International Archives of Photogrammetry and Remote Sensing, Bethesda, MD: ASPRS, Vol. XXIX, Comm. III, pp. 167-172
- Michaelsen, E., Soergel, U., Thoennessen, U. 2006. Perceptual grouping for automatic detection of man-made structures in high-resolution SAR data. *Pattern Recognition Letters*, 27(2), pp. 218-225.
- Michaelsen, E., Stilla, U., Soergel, U., Doktorski, L. 2010. Extraction of building polygons from SAR images: Grouping and decision-level in the GESTALT system. *Pattern Recognition Letters*, 31(10), pp. 1071-1076.
- Michaelsen, E., 2014. Gestalt algebra - a proposal for the formalization of gestalt perception and rendering. *Symmetry*, 6(3), pp. 566-577.
- Michaelsen, E., Yashina, V. V., 2014. Simple Gestalt Algebra. *Pattern Recognition and Image Analysis*, 24(4), pp. 542-551.
- Matsuyama, T., Hwang, V.S.-S., 1990. *SIGMA: A Knowledge-Based Image Understanding System*. Plenum Press, New York.
- Sassi, M., Demeyer, M., Wagemans, J. 2014. Peripheral Contour Grouping and Saccade Targeting: The Role of Mirror Symmetry. *Symmetry*, 6(1), pp. 1–22.
- Treder, M.S., 2010. Behind the Looking-Glass: A Review on Human Symmetry Perception. *Symmetry*, 2(4), pp. 1510–1543.
- Zhu, L., Chen, Y., Yuille, A. 2009. Unsupervised learning of probabilistic grammar-markov models for object categories, *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 29(31), pp. 114-128.