SEARCHING REMOTELY SENSED IMAGES FOR MEANINGFUL NESTED GESTALTEN

E. Michaelsen[a], D. Muench[a], M. Arens[a],

[a] Fraunhofer IOSB, 76275 Ettlingen, Germany - (eckart.michaelsen, david.muench, michael.arens)@iosb.fraunhofer.de

ICWG III/VII

KEY WORDS: Perceptual grouping, Symmetry, Urban structure recognition

ABSTRACT:

Even non-expert human observers sometimes still outperform automatic extraction of man-made objects from remotely sensed data. We conjecture that some of this remarkable capability can be explained by Gestalt mechanisms. Gestalt algebra gives a mathematical structure capturing such part-aggregate relations and the laws to form an aggregate called Gestalt. Primitive Gestalten are obtained from an input image and the space of all possible Gestalt algebra terms is searched for well-assessed instances. This can be a very challenging combinatorial effort. The contribution at hand gives some tools and structures unfolding a finite and comparably small subset of the possible combinations. Yet, the intended Gestalten still are contained and found with high probability and moderate efforts. Experiments are made with images obtained from a virtual globe system, and use the SIFT method for extraction of the primitive Gestalten. Comparison is made with manually extracted ground-truth Gestalten salient to human observers.

1. INTRODUCTION

Experts in gaining information from remotely sensed imagery, when performing e.g. map-update tasks on satellite images, use specific knowledge such as shadow directions, lay-over, and corner reflection phenomena in SAR, or specific spectral properties depending on the type of sensor. They are well aware of the specific type of knowledge they utilize.

On the other hand any non-expert human observer can still *see* meaningful part-and-aggregate structure, figure-background separation, and salient, i.e. probably meaningful, objects. Actually such observer may well achieve much better recognition performance as compared to our contemporary automatic methods. Such human beholder strongly relies on her/his capability of perceptual grouping according to Gestalt laws. Such grouping is very fast and occurs unknowingly to the subject. It also needs no specific training and works on any kind imagery including unfamiliar mapping modes such as SAR or thermal data. And more than that, since the expert is also a human being, she/he cannot 'switch off' Gestalt perception when looking at images, even if they were deceiving.

2. RELATED WORK

After recalling some approaches for automatic recognition of urban structure this section will concentrate on related work from Gestalt grouping, which is an interdisciplinary field between machine vision and psychology.

2.1 Automatic recognition of urban structure:

Several decades ago work on automatic understanding of urban aerial images started with knowledge-based approaches (Matsuyama and Hwang, 1990, Narasimham, 1970). Often, syntactic approaches or production systems were proposed. Also semantic representations such as semantic-nets were discussed. A good overview of such work a decade later can be found in the Ascona workshop volumes of the ETH (Baltsavias et al., 2001, Gruen et al., 1997, Gruen et al., 1995). More recently there are still some

syntactic approaches (Han and Zhu, 2005). Grammars also remain very popular in the field of facade understanding – see e.g. (Ripperda, 2008). Most often no real distinction is made between the automation of pre-attentive grouping and the automatic utilization of knowledge – the latter being a process of which the observer is usually quite aware.

2.2 Perceptual grouping using Gestalt principles:

Wertheimer's work may be regarded as initial starting point (Wertheimer, 1923) – though in the same period there were many important other names such as Metzger or Koffka, and some fundamental ideas are older and date back to Mach, Laplace, and Helmholtz.

Agnes Desolneux proposed a theory for machine perception of Gestalten based on statistical a-contrario tests (Desolneux et al., 2008). Two models – a uniformly distributed background and some foreground model (e.g. a Gaussian) – compete. If the primitives extracted from the image form an arrangement that is very unlikely with respect to the background model, the arrangement will be called an ϵ-meaningful Gestalt, where ϵ refers to the test level. She also introduces the notion of a *maximal ϵ-meaningful Gestalt*. It remains to be shown that the assessment functions defined below in Section 3. can be derived from such a-contrario testing. Intuitively, the same intention can be seen. Desolneux is aware that generic repetition of Gestalt formation on rising scales remains the goal, but up to now, to our knowledge, has not published any progress in that direction. She refers, however, to older challenges in that generic direction such as (Bienstock et al., 1997) and emphasizes that such ideas can even be found in very old statements of Helmholtz and Laplace.

3. GESTALT ALGEBRA

Standard reference to Gestalt algebra is (Michaelsen and Yashina, 2014) with more practical work being presented in (Michaelsen, 2014). It attempts capturing the laws of Gestalt-grouping in mathematical notations opening the way to machine generation as well as machine analysis - i.e. automatic rendering and recognition of Gestalten. Gestalt algebra is rather contrary to modern machine

learning; no example data are needed at all. Thus, the whole structure can be set up without any learning examples. It is not dependent on any data being representative or not. The following subsections briefly recall the basic definitions. First the domain is given on which the operations are defined, and then two operations are given namely for mirror symmetry and good continuation in rows. A third operation for rotational patterns introduced in (Michaelsen and Yashina, 2014) is not used in this work.

3.1 Gestalt domain

The domain G is defined as

$$ G \;=\; \mathbb{R}^2 \times \mathbb{R}^+ \times \mathbb{R}/\mathbb{Z} \times [0,1] \times \mathbb{N}. \tag{1} $$

Accordingly, for any Gestalt $g \in G$ there are **position** $po(g) \in \mathbb{R}^2$ (presented as center of a circle in the figures below), **scale** $sc(g) \in \mathbb{R}^+$ (presented as diameter), **orientation** $or(g) \in \mathbb{R}/\mathbb{Z}$ (presented as angle with respect to the horizontal axis), **assessment** $as(g) \in [0,1]$ and **frequency** $fr(g) \in \{1, 2, ...\}$ (presented as number of spokes in the circle) respectively. On the domain G operations are given. Algebraic closure demands that for any combination of Gestalten (parts) using any of the operations of the algebra, the newly constructed Gestalt (aggregate) is in fact again an element of the same domain. The key idea is that the operations yield maximal assessment value only for configurations of the parts that fulfill the intended Gestalt-laws perfectly. Random arrangements of parts have a low probability of causing high assessments. In fact most configurations lead to assessments close to zero. But the transition between these two extremes is continuous. Thus, Gestalt algebra is not a picture grammar because it does not set any hard geometrical constraints. Also no arbitrary constraints are allowed; instead very specific laws are used – the very laws of Gestalt perception.

3.2 Mirror symmetry

A binary operation is defined as $| : G \times G \to G$, and the terms are given in infix notation as $g = h_1 | h_2$. The components of g result as follows:

$$ po(g) \;=\; \tfrac{1}{2}\left(po(h_1) + po(h_2)\right) \tag{2} $$

$$ sc(g) \;=\; |po(h_2) - po(h_1)| + \sqrt{sc(h_1) \cdot sc(h_2)} \tag{3} $$

$$ or(g) \;=\; \tfrac{1}{\pi}\arctan\left(\frac{po_x(h_2) - po_x(h_1)}{po_y(h_2) - po_y(h_1)}\right) \tag{4} $$

$$ as(g) \;=\; \sqrt[4]{a_{|p} \cdot a_{|s} \cdot a_{|o} \cdot a_{|a}} \tag{5} $$

$$ fr(g) \;=\; 2 \tag{6} $$

where the function $\arctan$ in Equation (4) is set to $\pi/2$ if zero appears in denominator. The components of the geometric mean in Equation (5) code the intended Gestalt-principle of mirror symmetry. They are given as follows:

$$ a_{|p} \;=\; \sqrt{e}\frac{|po(h_2) - po(h_1)|}{\sqrt{sc(h_1) \cdot sc(h_2)}}e^{-\frac{1}{2}\frac{|po(h_2) - po(h_1)|^2}{sc(h_1) \cdot sc(h_2)}} \tag{7} $$

$$ a_{|s} \;=\; e^{2 - \frac{sc(h_1)}{sc(h_2)} - \frac{sc(h_2)}{sc(h_1)}} \tag{8} $$

$$ a_{|o} \;=\; \begin{cases} 0, & \text{if } fr(h_1) \neq fr(h_2) \\ \frac{1}{2} + \frac{1}{2}\cos(\beta), & \text{else.} \end{cases} \tag{9} $$

$$ \beta \;=\; 2\pi\left((or(h_1) + or(h_2)) - fr(h_1)(or(g) + \tfrac{1}{2})\right) $$

$$ a_{|a} \;=\; \sqrt{as(h_1) \cdot as(h_2)} \tag{10} $$

In detail: $a_{|p}$ prefers configurations where the parts are not too far away and not too close to each other (as compared to their scale);

this differs a little from the definition (4) given by (Michaelsen and Yashina, 2014) having the shape of a Rayleigh distribution here – and thus less mass on the ratios greater than one; $a_{|s}$ prefers equal scales of the parts; $a_{|o}$ prefers mirror symmetric orientations with respect to the perpendicular bisector axis; and $a_{|a}$ inherits the assessments of the parts.

3.3 Good continuation in rows

An n-ary operation is defined as $\Sigma : G^n \to G$, and the terms are notated in prefix notation as $g = \sum h_i$ where $1 \leq i \leq n$. The components of g result as follows:

$$ po(g) \;=\; \tfrac{1}{n}\sum_{i=1}^{n} po(h_i) \tag{11} $$

$$ sc(g) \;=\; |po(h_n) - po(h_1)| + \sqrt[n]{sc(h_1)\cdots sc(h_n)} \tag{12} $$

$$ or(g) \;=\; \tfrac{1}{\pi}\arctan\left(\frac{po_x(h_n) - po_x(h_1)}{po_y(h_n) - po_y(h_1)}\right) \tag{13} $$

$$ as(g) \;=\; \sqrt[4]{a_{\Sigma p} \cdot a_{\Sigma s} \cdot a_{\Sigma o} \cdot a_{\Sigma a}} \tag{14} $$

$$ fr(g) \;=\; 2 \tag{15} $$

For simplicity Equation (13) sets the orientation of the new Gestalt from the first and last part only. Another – and more adequate – way would be to set it by regression over all n parts. The components of the geometric mean in Equation (14) code the intended Gestalt-principle of good continuation in a row. They are given as follows:

$$ a_{\Sigma p} \;=\; \sqrt{e}\frac{|po(h_n) - po(h_1)|}{sc_{mid}(n-1)}e^{-\frac{1}{2}\frac{|po(h_n) - po(h_1)|^2}{sc_{mid}(n-1)^2}} \tag{16} $$

$$ \cdot \left[\prod_{i=1}^{n}\exp\left(-\frac{|po(h_i) - set_i|}{sc_{mid}}\right)\right]^{\frac{1}{n-2}} \text{ with} $$

$$ set_i \;=\; po(g) + \left(\frac{i-1}{n-1} - \frac{1}{2}\right)(po(h_n) - po(h_1)) \tag{17} $$

$$ a_{\Sigma s} \;=\; e^{2n - \frac{sc(h_1)}{sc_{mid}} - \frac{sc_{mid}}{sc(h_1)} - \cdots - \frac{sc(h_n)}{sc_{mid}} - \frac{sc_{mid}}{sc(h_n)}} \tag{18} $$

$$ a_{\Sigma o} \;=\; \left(\prod_{i=1}^{n}\left(\frac{1}{2} + \frac{1}{2}\cos(\beta_i)\right)\right)^{1/n} \text{ with} $$

$$ \beta_i \;=\; 2\pi(or(h_i) - ao(h_1\cdots h_n)) \tag{19} $$

$$ a_{\Sigma a} \;=\; (as(h_1)\cdots as(h_n))^{1/n} \tag{20} $$

In detail: $a_{\Sigma p}$ prefers configurations where the parts are neatly aligned in a row. To this end the ideal positions set_i of such a row must be obtained by Equation (17). It also prefers the mid-scale to be about the same as the spacing (again by use of a Rayleigh distribution form), $a_{\Sigma s}$ prefers equal scales of the parts. The geometric mean scale sc_{mid} of the parts must be calculated for use in Equation (16) as well as in Equation (18). In Equation (19) $a_{\Sigma o}$ prefers orientations close to the average orientation ao which is calculated (using complex numbers) by

$$ ao(h_1\cdots h_n) \;=\; \tfrac{1}{2\pi}\arg\left(\sum_{i=1}^{n}\exp(2\pi\, or(h_i))\right) \tag{21} $$

It makes only sense if the frequencies of all parts are equal. Else $a_{\Sigma o} = 0$ is set just like in Equation (9). Finally, $a_{\Sigma a}$ inherits again the assessments of the parts.

4. RECURSIVE SEARCH FOR GESTALTEN

Given a finite set of primitive Gestalten $P \subset G$ and a recognition level $0 \leq \epsilon < 1$ the recognition task can be formulated as:

Find all $g \in G$ with $as(g) \geq 1 - \epsilon$ such that $g = t$ where t is a Gestalt algebra term using only Gestalten from P and the basic operations $|$, and Σ.

One obvious algorithmic solution to this task is a recursive enumeration of all terms, and testing them for the property $as(g) \geq 1 - \epsilon$. At first glance this enumeration might never stop because in algebraic terms one $p \in P$ may appear multiply. In (Michaelsen and Yashina, 2014) Lemmas are given stating that repetition will lead to arbitrary small assessments. And so the main proposition below follows:

Main Proposition: For any $1 \geq \epsilon > 0$ the set of terms better than $1 - \epsilon$ defined by the nested recursion of Algorithms 1 to 4 on a finite set of primitives is *finite*.

This means that the recursive search surly ends provided there are enough computational resources. The recursively nested set given below unfolds again only a sub-set of the possible Gestalten: 1) Usually it will not be useful to list all partial rows of better assessment than $1 - \epsilon$ of a given long $\sum$-Gestalt. In (Desolneux et al., 2008) the notion of *maximal meaningful* row-Gestalten is introduced stating that such rows should be prolonged by adding more and more parts from the basis-set until a certain evaluation – in the case of the a-contrario test theory of Desolneux the false alarm rate FAR, in the case of Gestalt-algebra the assessment – starts getting worse. Finding this point is the goal of the following Algorithm 1. Recall that the algorithm works on sets, so that the set will not become bigger, if an element is added which is already present (because it has been constructed by applying a different sequence of calls on the same parts). 2) The handling of lists below ensures that only terms of the same operation are combined. E.g. no mirror-symmetry Gestalt will be formed of a row-Gestalt on the one hand and a mirror-symmetry on the other hand. Information on the decomposition is transported from one part to the others. Again such search only unfolds a very small subset of the – already finite – sub set of all terms better than $1 - \epsilon$. The point of this contribution is that the really meaningful Gestalten are still contained. 3) Obviously, the recursion does not go deeper when the scale of the Gestalten is larger then the e.g. input image.

Algorithm 1: main

Input: *BasicGestalten*
Output: sorted list of recognized *Gestalten*

1 **begin**
2   sort *BasicGestalten* descending concerning each assessment $as(g)$;
3   *BasicGestalten* ← select Gestalten of *BasicGestalten* with $as(g) < \epsilon$;
    `// recognize initial |-Gestalten.`
4   *ListMirror* ← calculateMirror(*BasicGestalten*,ϵ_{mirror});
    `// recognize initial Σ-Gestalten consisting`
    `of two primitives.`
5   *ListRow* ← calculate2Row(*BasicGestalten*,ϵ_{row});
    `// computing recursively Σ-Gestalten.`
6   *ListRow* ← calculateRow(*BasicGestalten*, *ListRow*, ϵ_{row});
    `// functional recursion`
7   *Gestalten* ← recursiveGestalting(*ListMirror*, *ListRow*);
8   sort *Gestalten* descending concerning each assessment $as(g)$;
9   **return** *Gestalten*;

Algorithm 2: calculateMirror (Searching for maximal meaningful mirrors)

Input: *BasicGestalten*,ϵ_{mirror}
Output: *ListMirror*

1 **begin**
2   **forall** $g_m \in BasicGestalten$ **do**
3   **forall** $g_{m_2} \in BasicGestalten$ **do**
4             $g_{m_{new}} \leftarrow computeSymmetry(g_m, g_{m_2})$;
5   **if** $as(g_{m_{new}}) < \epsilon_{mirror}$ **then**
6   *ListMirror* ← *ListMirror* + $g_{m_{new}}$;
7   **return** *ListMirror*;

Algorithm 3: calculateRow (Searching max. meaningful rows)

Input: *BasicGestalten, ExtRowGestalten*,ϵ_{row}
Output: *ExtRowGestalten*

1 **begin**
2   **forall** $g_r \in ExtRowGestalten$ **do**
3   **forall** $g_{r_b} \in BasicGestalten$ **do**
4             $g_{r_{new}} \leftarrow computeRow(g_r, g_{r_b})$;
5   **if** $as(g_{r_{new}}) < \epsilon_{row}$ **then**
6   *ExtRowGestalten* ← *ExtRowGestalten* + $g_{r_{new}}$;
7   **if** *ExtRowGestalten* $= \emptyset$ **then**
8   **return** $\emptyset$;
9   **else**
10   **return** *ExtRowGestalten* ← calculateRow(*BasicGestalten, ExtRowGestalten,* ϵ_{row});

Algorithm 4: recursiveGestalting (Recursive search for Gestalten)

Input: *ListMirror, ListRow*
Output: *Gestalten*

1 *hasNewGestalten* ← false;
  `// generate new |-Gestalten`
2 *ListMirrorRec* ← {};
3 *ListMirrorRec* ← calculateMirror(*ListMirror*,ϵ_{mirror});
4 *ListMirrorRec* ← *ListMirrorRec* + calculateMirror(*ListRow*,ϵ_{mirror});
5 **if** *ListMirrorRec* $\neq \emptyset$ **then**
6   *hasNewGestalten* ← true;
  `// generate new Σ-Gestalten`
7 *ListRowRec* ← {};
8 *ListRowRec* ← calculateRow(*ListMirror*,*ListRow*,ϵ_{row});
9 *ListRowRec* ← *ListRowRec* + calculateRow(*ListRow*,*ListRow*,ϵ_{row});
10 **if** *ListRowRec* $\neq \emptyset$ **then**
11   *hasNewGestalten* ← true;
12 **if** *hasNewGestalten* **then**
    `// functional recursion goes even deeper`
13   *Gestalten* ← {};
14   *Gestalten* ← recursiveGestalting(*ListMirrorRec*,*ListRowRec*);
15   *Gestalten* ← *Gestalten* + *ListMirrorRec* + *ListRowRec*;
16   **return** *Gestalten*;
17 **else**
    `// Recursion termination. Returns empty list.`
18   **return** $\emptyset$;

Table 1: Remotely sensed imagery dataset on which the proposed method is evaluated. In the right three columns the results are quantitatively listed.

#	North			East			Height	False N	False P	True P
1	33°	28'	14.32''	-112°	02'	03.80''	500m	4	0	0
2	33°	28'	00.42''	-112°	02'	24.27''	500m	4	3	0
3	33°	27'	16.73''	-112°	01'	38.61''	500m	1	13	4
4	33°	27'	20.28''	-112°	02'	28.48''	500m	3	5	0
5	33°	37'	01.86''	-112°	17'	23.56''	500m	6	26	11
6	33°	27'	25.04''	-112°	02'	25.54''	2000m	2	1	1
7	33°	36'	56.11''	-112°	17'	05.00''	2000m	3	240	0
8	33°	28'	49.77''	-112°	03'	18.06''	8000m	5	0	0
9	27°	28'	22.90''	89°	38'	15.65''	500m	1	0	0
10	27°	27'	52.42''	89°	38'	28.88''	500m	8	2	0
11	27°	26'	30.81''	89°	39'	59.21''	500m	2	0	0
12	27°	27'	16.06''	89°	39'	17.74''	500m	0	7	2
13	27°	28'	40.91''	89°	37'	22.24''	500m	4	1	10
14	27°	28'	18.72''	89°	37'	51.12''	2000m	0	0	0
15	48°	56'	18.66''	8°	57'	07.97''	500m	3	2	0
16	48°	56'	17.38''	8°	57'	34.28''	500m	2	27	0
17	48°	55'	51.00''	8°	57'	36.05''	500m	3	5	0
18	48°	55'	32.53''	8°	57'	33.52''	500m	0	0	0
19	48°	55'	39.09''	8°	58'	05.41''	500m	0	0	0
20	48°	56'	00.65''	8°	57'	45.30''	2000m	0	0	0

5. EXPERIMENTS ON REMOTELY SENSED IMAGERY

In order to demonstrate robustness with respect to geographic location three cities on three different continents were chosen: Thimphu (Bhutan), Vaihingen (Germany), and Phoenix (Arizona). The images where obtained from salient structures there using the Google Earth virtual globe system - see Table 1 for the geo location and Figure 1 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.

The images as given by the virtual globe system where cropped so as to remove the logos for the automatic processing, and only the intensity is used - not the colors. Preliminarily, the standard SIFT-key-point detector (SIFT++) was used to extract a set of primitive Gestalten from the intensity images. The SIFT-key-point detector fits the Gestalt domain, because it directly gives position, orientation, scale, and assessment. The frequency attribute of all primitives was set to one.
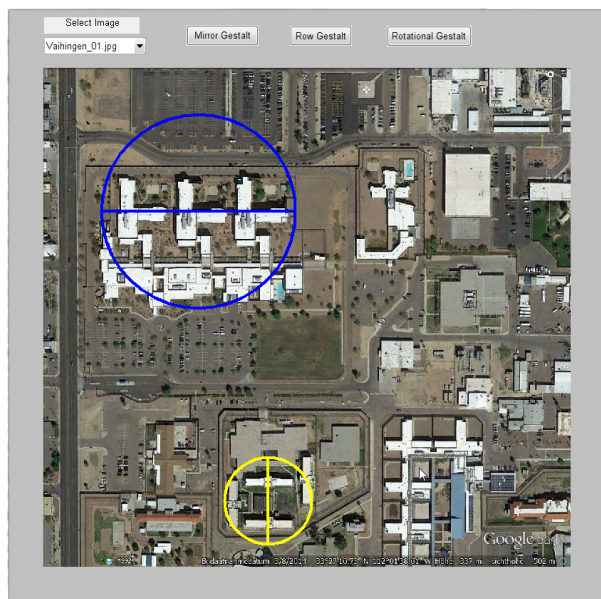
5.1 Ground truth

A graphical user interface (GUI) was constructed allowing human observers to mark salient Gestalten on images in accordance with the definition of the Gestalt domain in Equation 1. The observer may mark one or more mirror Gestalten or row Gestalten on the image by mouse operations. He or she may also decide not to mark anything, if nothing salient should appear in the image. Figure 1(a) shows an example.

5.2 Results

The automatic search outlined above yields Gestalt algebra terms of arbitrary nesting depth, while the ground truth given by the GUI described above will only consist of a – possibly empty – set of Gestalten. For the time being we do not expect very frequent accordance in the part-of tree structure a human would prefer with the corresponding term structure found by the automatic search. It would also be a question on its own, how to extend the GUI in such a way that the observer can construct Gestalt hierarchies, without being too complicated for non-expert users. As mentioned before, Gestalt perception is unknowingly and very swiftly performed, and it is a non-expert capability. So for the time being we only compare the root node of the set of Gestalten automatically found to be above a suitable threshold with the set of salient Gestalten obtained via the GUI.

Gestalten will be regarded as matching if the position fits up to 25% of the geometric average of both scales, and if the orientation fits up to 30°. The last three columns in Table 1 list the numbers of false negatives, false positives, and true positives for each picture and thus gives a quantitative estimate on the currently achieved accordance with human perception.

(a)
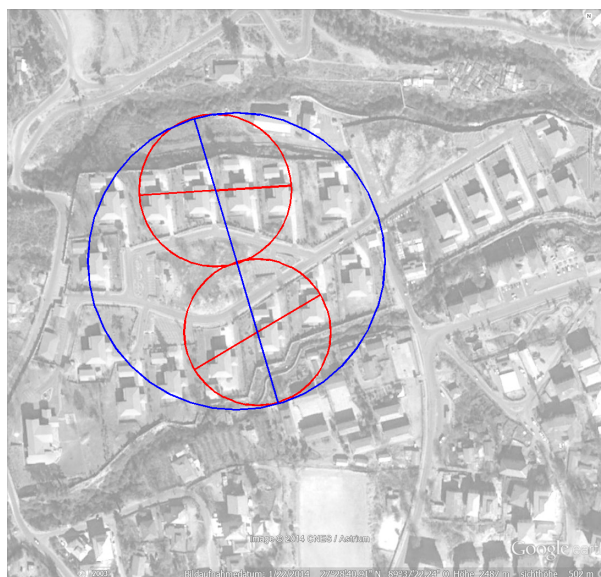
(b)

Figure 1: (a) GUI for ground-truth acquisition: Salient row-Gestalt in blue, salient mirror-Gestalt in yellow. (b) Some result on image #11: Salient mirror-symmetry-Gestalt in blue, it's part row-Gestalten in red.

6. CONCLUSION AND DISCUSSION

Gestalt algebra can be useful in almost any machine recognition task from pictorial data. Usually it will be applied as an intermediate step before domain specific semantics are appropriate. Neither representative learning data nor domain specific knowledge is required. In this contribution it is shown that recursive search for such Gestalten from a finite set of primitives given with an

image is possible, as long as only objects of a certain minimal quality count.

Furthermore, again only a subset is of interest, where structural information is transported from one part inside an aggregate to the other parts. E.g. starting from the ten small Gestalten displayed in Figure 1 (b) the large central mirror-symmetry Gestalt will be found by the search outlined in Section 4., because both parts of the intermediate level are in fact row Gestalten. The assessment of the final mirror-symmetry Gestalt is not getting smaller because the two rows have a different number of members. This is not part of our automatic recognition yet. Thus, the search yields still a larger set than intended and formalisms remain to be given for such kind of non-local information transport. Other future work will consider different primitive extraction methods, such as e.g. super-pixels, and/or the use of spectral information as additional source of similarity or dissimilarity.

ACKNOWLEDGEMENTS

REFERENCES

Baltsavias, E.-P., Gruen, A. and Van Gool, L., 2001. Automatic Extraction of Man-made Objects from Aerial and Space Images (III). Balkema Publishers, Lisse.

Bienstock, E., Geman, S. and Potter, D., 1997. Compositionality, MDL Priors, and Object Recognition. MIT Press, chapter 9, pp. 838–844.

Desolneux, A., Moisan, L. and Morel, J.-M., 2008. From Gestalt Theory to Image Analysis. Springer, Berlin.

Gruen, A., Baltsavias, E.-P. and Henricsson, O., 1997. Automatic Extraction of Man-made Objects from Aerial and Space Images (II). Birkhaeuser, Basel.

Gruen, A., Kuebler, O. and Agouris, P., 1995. Automatic Extraction of Man-made Objects from Aerial and Space Images. Birkhaeuser, Basel.

Han, F. and Zhu, S.-C., 2005. Bottom-up/top-down image parsing by attribute graph grammar. In: ICCV, IEEE Computer Society, pp. 1778–1785.

Matsuyama, T. and Hwang, V. S.-S., 1990. SIGMA: A Knowledge-Based Image Understanding System. Plenum Press, New York.

Michaelsen, E., 2014. Gestalt algebra - a proposal for the formalization of gestalt perception and rendering. Symmetry 6(3), pp. 566–577.

Michaelsen, E. and Yashina, V. V., 2014. Simple gestalt algebra. Pattern Recognition and Image Analysis 24(4), pp. 542–551.

Narasimham, R., 1970. Picture languages. In: Picture Language Machines, Conference in Canberra 1969, Editor: Kaneff, S., Academic Press, London, pp. 1–30.

Ripperda, N., 2008. Grammar based facade reconstruction using rjmcmc. Photogrammetrie Fernerkundung Geoinformation 2, pp. 83–92.

Wertheimer, M., 1923. Untersuchungen zur lehre der gestalt, ii. Psychologische Forschung 4, pp. 301–350.