

ON THE AUTOMATION OF GESTALT PERCEPTION IN REMOTELY SENSED DATA

E. Michaelsen¹¹Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Gutleuthausstr. 1, 76275 Ettlingen, Germany**Abstract**

Gestalt perception, the laws of seeing, and perceptual grouping is rarely addressed in the context of remotely sensed imagery. The paper at hand reviews the corresponding state as well in machine vision as in remote sensing, in particular concerning urban areas. Automatic methods can be separated into three types: 1) knowledge-based inference, which needs machine-readable knowledge, 2) automatic learning methods, which require labeled or un-labeled example images, and 3) perceptual grouping along the lines of the laws of seeing, which should be pre-coded and should work on any kind of imagery, but in particular on urban aerial or satellite data. Perceptual grouping of parts into aggregates is a combinatorial problem. Exhaustive enumeration of all combinations is intractable. The paper at hand presents a constant-false-alarm-rate search rationale. An open problem is the choice of the extraction method for the primitive objects to start with. Here super-pixel-segmentation is used.

Keywords: perceptual grouping, remote sensing, urban areas.

Citation: Michaelsen E. On the automation of gestalt perception in remotely sensed data. *Computer Optics* 2018; 42(6): 1008-1014. DOI: 10.18287/2412-6179-2018-42-6-1008-1014.

Acknowledgments: Google Earth provides remotely sensed imagery for almost any location on the planet freely to public use. I picked some of these images as example data as well for explanation of the issue as for experiments with our machine vision approach. I thank Google for the data and acknowledge the substantial contribution that Google gives to the scientific community with such data.

Introduction

Gestalt grouping or perceptual grouping of certain forms of repetitions and regularities has long ago been identified as important mechanism of seeing. Long before the introduction of computing machines, psychologists have gathered substantial evidence for its important role, the most popular reference being *Wertheimer* [8]. However, many other important figures have been contributing such as *Mach* or *Helmholtz*, and one of the most recommendable contemporary textbooks in this subject by *Pizlo et al.* even cites a thousand year old work on the very topic [4].

Contemporary interdisciplinary approaches, such as the one given by *Pizlo's* group, start from evidence gathered by experiments conducted with a set of human observers looking on computer generated graphics. Then the work continues by coding machines in such way as to reproduce the same grouping phenomena that were found by the observer trials as good as possible. In the end, the performance of the seeing machines is compared with other approaches to machine vision, or evaluated in a perception-action loop with a robot in a more or less natural environment.

Almost like a singularity in the continuum of Gestalt literature stands out the book of *Leyton* [5]. It states that perception is always inference from detectable traces on untouched homogenous ground, i.e. symmetry-breaking distortions on symmetric background that does not give any information on its own. This is indeed a counter-intuitive argument – not the symmetric arrangement is the foreground, instead we perceive scratches on a wall and infer that something must have been moving along it. Seeing is understood as inversion of causality and processes.

When studied without prejudice and with diligence an algebraic spirit appears in this work, really treating the generative core, the hierarchies in which certain Gestalten are encountered in visual data, through the scales. Ac-

ordingly, *Leyton* develops specific own notations using operators and bracketing.

Based on *Grenander's* and *Mumford's* algebraic approach to machine vision, where the constructive parts-to-aggregate nature of seeing was most strongly emphasized, *Desolneux* presented a mathematical theory for the use of Gestalt laws in machine vision [1, 3]. “Mathematical” refers here to probabilistic: The detection of a (symmetrical) foreground Gestalt aggregate on more or less asymmetric – i.e., uniformly distributed – background clutter is founded on statistical tests. Unfortunately, two statistical models are required – usually a uniform background and a normally distributed deviation from the Gestalt laws in the foreground. The corresponding two-model approaches to statistical testing are known as a contrario testing – a technical and philosophical difficult combinatorial subject. Probably due to these technical problems, the approach lacks generative depth in hierarchies of patterns through the scales.

Liu and her group from Pennsylvania State University organized symmetry competitions along with major machine vision conferences [11, 12]. In the corresponding public data with ground truth numerous faces, vehicles, and animals can be found along with some facades and architecture, but no remotely sensed imagery. Following the major trends in machine vision, *Liu* and her group switched to deep learning machine vision recently.

Decomposition of remotely sensed images in a generative hierarchy of parts and aggregates has been studied as knowledge-based artificial intelligence topic in the last century in events like the Ascona workshops of the ETH Zurich [6–8] and in the image understanding workshops of the DARPA [9]. The prevailing approach then being knowledge-based analysis, e.g. by semantic nets, systems of production rules, expert systems, and automatic inference.

Our own approach comprises properties of all of these approaches. Starting from knowledge-based automatic reasoning on a set of primitive objects extracted from an aerial image, we were mainly interested in building and road recognition [13]. Then such recognition method was integrated in a perception-action loop for unmanned aerial vehicle navigation [14]. One main advantage of such knowledge-based inference over more conventional machine learning approaches – such as convolutional or deep-learning neural nets – is their independence of example data. Knowledge-based methods can be applied directly to the incoming pictures, while training all appearances of objects like bridges or churches with a machine-learning method requires a huge set of representative and carefully labeled data to be processed prior to application. Moreover, the recognition performance will depend crucially on the learning data being representative for the data occurring later in the application.

On the other hand, knowledge-based approaches require the loading of machine-readable knowledge prior to application. That can be a problem as well, because editing knowledge is laborious and error prone, and by the time of execution the rules may be outdated or semantically not appropriate at all.

Many of the rules utilized during that work on urban remote sensing data turned out to be similar, no matter whether aerial or satellite images, thermal data, multi- or hyperspectral data, or even synthetic aperture radar were used [15]. Pairs of reflection symmetric parts were important over and over again, as well as repetition of parts in rows, and prolongation in good continuation along linear structures, including gap filling. We found that these rules and aggregations where in fact not domain-specific knowledge. Instead they are ubiquitous, can be coded only once for all application, and generalize perfectly to any new task. They are the same rules, that are known to psychologists as the laws of seeing, the Gestalt perception, with which nature equips humans and animals, so that they are fit for encountering their environment prior to any learning.

Accordingly, we proposed a set of operations operating on a fairly general image domain – the Gestalt algebra [17].

Motivation

Fig. 1 displays a remotely sensed example image of an urban region. The location was picked rather arbitrarily (guided by the location of the ITNT conference) and not on purpose to pick a biased example with particular structure. Yet, a lot of Gestalt seeing issues can be discussed also on this example – just as on almost any other urban image:

The most salient object in this example attracting instantaneously any observer's attention is a strange black and white pattern on an elliptic building to the Northeast. "Pattern" must be understood in its common sense meaning here, not in its technical use in machine learning. Obviously, a lot of algebra rules the construction and appearance of such objects. Maybe the designing architect was a fan of *Escher* and *Kanizsa*. Parts are repeated using certain transforms that are members of algebraic entities.

The architects – i.e. the generating side – always have an advantage over the coders of automatic understanding of remotely sensed data – i.e., the analyzing side. Architects can choose whatever they want among the algebraic structures, and if they like *Fibonacci*, they can use that, why not! The analyzing side will probably be surprised by such structure.

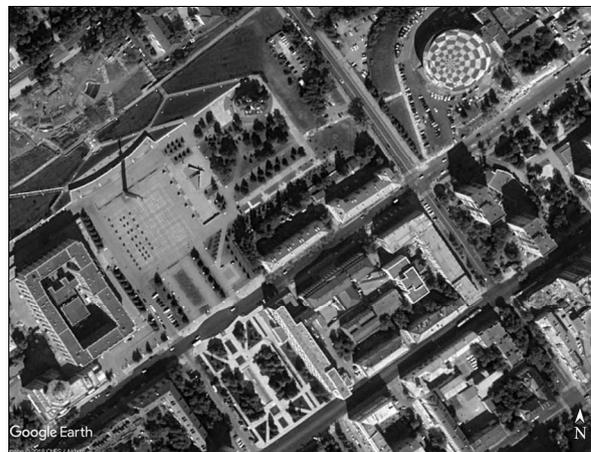


Fig. 1. A more or less arbitrary urban satellite image, courtesy to Google Earth

However, one thing remains evident: There will be some algebraic structure present in the data. In the example, symmetries are not only given in that most salient building. For instance, there is a strong multiple reflection symmetry in the park to the South East. Rectangularity and parallelism are preferred in the road grid, similar architectural arrangements in groups of buildings to the East form a hierarchy. Trees come in rows, etc.

While all these deep and complex hierarchies are evidently present in the 3D scene, much is lost in the projection into the 2D image domain. In the symmetry recognition literature, this is regarded as the main problem. *Pizlo* states, that symmetry is almost never preserved when viewed through a camera or an eye [4], almost never meaning with probability zero. He calls such cases, where the symmetry survives, *degenerate*. However, in remote sensing the situation is much more benign. Satellite images, such as the one in Figure 1, do not change the symmetric arrangements in the ground-plane.

Three serious problems remain: 1) lighting can break the reflection symmetry of gabled roofs, as we can see in the example; 2) shadows cast by high objects disturb the appearance of lower objects, and 3) partial occlusion due slight oblique view directions.

The gestalt domain and some operations on it

Any element considered in such reasoning will feature a *location* in the image. That can be pixel coordinates in row and column, or geographic coordinates in North and East. We neglect image margins and pixels raster, as well as any Earth curvature, and assume the 2D plane for this location feature, because of the many algebraic advantages such vector space provides. Any element also features a *scale*. Algebra teaches that the scales form a multiplicative and continuous group of

dimension one, bounded by zero on the small side and with no limit on large scales. Almost any element also features an *orientation*. For simplicity, we demand it for all, being meaningless in rare exceptional cases of complete rotary symmetry. Orientations are an additive and continuous group.

While these features are intuitively evident, two other compulsory features of the Gestalt domain used in [17] may require more motivation: 1) *Rotational frequency* is a positive integer number n . It means that the object is self-similar under rotation of $2\pi/n$. Accordingly, for an object with $n = 2$ the orientation is represented by an angle only between 0 and π . Such additional frequency became necessary when hierarchical aggregates of high symmetry are considered. This is clear for aggregates with rotational symmetry. In addition, a row of similar parts remains the same, whether it is enumerated forward or backward, so that its rotational self-similarity frequency is 2. 2) *Assessment* turns out the most important feature of any object in the domain under consideration. This is a real value between *zero* (for meaningless) and *one* (most salient).

Intuition would understand the laws of seeing as hard constraints. A kind of picture grammar would use such constraints in its productions, building e.g. a mirror-aggregate of two parts, provided the corresponding geometric constraint was fulfilled. Pictures could then be generated or parsed in hierarchies of arbitrary depth, defining a picture language. The performance of such systems on deeper hierarchies turned out bad in two aspects: 1) in the reducing or parsing direction frequently one of many constraint thresholds is violated, so that recognition fails. And if the thresholds are chosen more liberal so that present hierarchies visible to human subjects can be reproduced, 2) the combinatorics of the search dictate enormous efforts so that it simply doesn't work.

Consequently, we decided to replace the hard constraints by *smooth assessment functions* – similar to the membership functions used in *Zadeh's* fuzzy logic. We prefer functions that are differentiable everywhere, following the shape of known distribution models such as *Rayleigh* or *Fisher*. Note, the term grammar or production rule is no longer appropriate in such setting. Instead, we speak of an *algebra* with *operations* on it. Any element of the domain can be combined with any other element, yielding always a new resulting element. If the configuration of the parts does not fulfill the corresponding Gestalt law at all, the resulting aggregate will be assessed as zero. It is meaningless, but it exists, and thus we have algebraic closure. Other frequent algebraic properties such as associativity or existence of neutral or inverse elements are not given here.

For remote sensing, the most important operations are one for reflection symmetry (a binary and commutative operation noted with $|$), and one for the formation of rows, which are also called frieze symmetries e.g. in the literature, such as [11] and [12]. This n -ary operation is noted by Σ , and commutative in a generalized sense (recall, n can be larger than two, and remains unspecified). We published variants of these operations e.g. in [16] and [17], and their application to remotely sensed data e.g. in [18] and [19].

The problem of primitive extraction

Up to now, in almost all our applications of Gestalt algebra we used the well-known scale-invariant feature-transform (SIFT) key-points for the extraction of the terminals or primitives, from which the search for Gestalten starts. Such key-points naturally give all the features required for our Gestalt domain, including a meaningful assessment. The rotational frequency was fixed to one.

In this paper, we introduce the utilization of super-pixel segmentation following [1] for this purpose. Figure 2 shows an example. That image has been used before in [18], where also the Geographic details are given. In remote sensing images such segments are located well on the objects of interest, while the SIFT key-points tend to mark corners on the contours of the objects. Thus, super-pixel primitives lead to results more consistent with human perception. The figure shows the procedure: Super-pixel segmentation is applied to the image using some default parameterization; very small and isolated fuzzy regions are removed (with the corresponding locations remaining black in Figure 2b); thus from about one million color pixels in a grid a set of about thousand segments is reckoned each having one average color; each super-pixel segment gives a location (its mean position), a scale (from the root of the number of pixels in it), and second moments, which are transformed into an orientation and an eccentricity. The figure only displays intensities, while the original has the usual three colors.

If the margins of the super pixel are enforced by the hexagonal grid (because it is located in a rather homogeneous region) it will be assessed bad. On the other hand, if it has margins resulting from strong contrast to its neighbors, it will be assessed well. Thus, all the compulsory features for the Gestalt domain are given. The rotational frequency for these primitives fixed to two, since they are invariant under rotation of π .

In addition to these features, we have colors, and an eccentricity with each primitive, providing extra information. In Figure 2c these objects are displayed as elliptic patches. Such graphical conventions for displaying sets of Gestalten are very important. It allows using the human Gestalt perception again, and assessing the loss of content when reducing the information, in this case caused by the primitive extraction method. Figures 2b and 2c almost look like pieces of art.

Recall, Thimphu is known among urban settlement experts as model city, concerning sustainable and intelligent urban planning. Figure 2b might be similar to a painting brushed by the designer of the place, capturing his composition in an expressive way. Figure 2c seems less felicitous, a little too abstract.

Many other primitive extraction methods are possible, and should be tested, from very simple ones – such as simple thresholding and using the connected components possibly subject to some morphological operations, to very sophisticated ones – such as using state-of-the-art semantic segmentation methods from the deep-learning community.

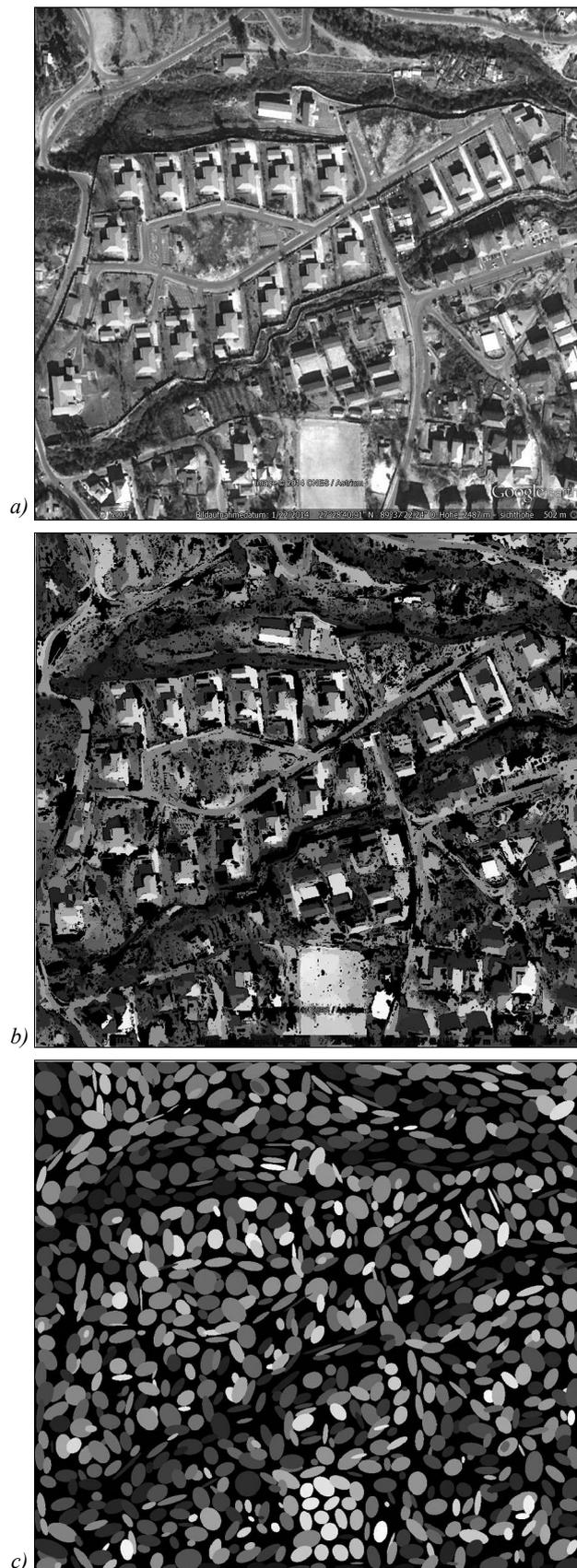


Fig. 2. Example image #5 – Thimphu, Bhutan, a) original, converted to intensities, b) super-pixel segments, c) super-pixel features (without colors), courtesy to Google Earth

Search

The combinatorial problem resulting from such generative models was already seen by the pioneers of the field such as *Rosenfeld* or *Grenander* clearly. While the parsing of strings, using grammars, such as the Chomsky types, remains tractable (provided context free models are used), parsing of pictures is generally intractable.

The closure of a given finite set of such Gestalt-operations, and a given finite set of primitives (the constants in our algebra) is even infinite, and can never be listed exhaustively. However, we proved that the set of corresponding Gestalten, assessed better than any given $\epsilon > 0$, is indeed finite and can be listed [17]. Still completely correct listing remains intractable. The number of possible aggregates can rise very dramatically with rising depth of hierarchy.

The easiest feasible way is following a *constant false alarm rate* approach: Only the best Gestalten of a certain level in the part-of hierarchy are accepted. How many depends on the computational resources and time available. Then only from these, the next level is constructed using the productions, and again only the best are accepted. Returning to the example image presented in Figure 2, we first list the sets of the first level of grouping, i.e.

$$L_1 = \{f = g \mid h; g, h \in P\}, \tag{1}$$

which is the set of all reflection symmetry Gestalten made of primitives and

$$L_\Sigma = \{f = \sum g_1 \cdots g_n; g_1 \cdots g_n \in P\}, \tag{2}$$

which is the set of all row-Gestalten made of primitives. P is the set of primitives depicted in Figure 2c.

Obviously, the enumeration in (1) is of quadratic complexity in the number of primitives. However, we are not interested in meaningless badly assessed pairs. Thus, we can accept only the – say one hundred – best assessed aggregates in this set, bounding its size. Let us call this set L_{100} .

For larger images, quadratic complexity is not acceptable. However, one of the Gestalt laws used in all operations is *proximity*. Thus the distance, as compared to the mean scale of the parts, can be limited and only pairs closer to each other than that limit need to be tested. Thus, sub-quadratic computational complexity is possible for that step. There are also sub-quadratic methods for picking the best few – that task has a little lower complexity than sorting.

n -ary operations, such as \sum in (2), are a harder challenge concerning the search efforts. (2) defines the enumeration of all strings from elements of P . This is clearly of exponential complexity. The sub-set of meaningful rows, i.e. those with assessment not very close to zero, is very small and thin in that combinatorial set. One can start with listing pairs, just like for the binary operation in (1). These pairs are subsequently used as seeds. We showed that there is a greedy linear search method for such row-prolongation performing reasonable well [18]. From the resulting set of row-Gestalten, again only the – say one hundred – best assessed will be accepted, bounding its size. Let us call this set $L_{100\Sigma}$.

This was only the first level. Now, hierarchical aggregates can be constructed. For instance the set of reflection symmetric pairs of rows:

$$L_{|100\Sigma} = \{f = g \mid h; g, h \in L_{100\Sigma}\}. \quad (3)$$

Next to this $|\Sigma$ -combination, the second level contains also the other combinations $||$, $\Sigma|$, and $\Sigma\Sigma$, respectively. We know from the theorem in [17] that there is a bound for the depth in level. All sets of higher levels will be empty. Thus, the constant false alarm rate approach yields predictable and tractable search efforts.

In Figure 3 such search is exemplarily displayed. Part a) gives the primitives in the Gestalt-display convention, as used in [16–18], and [19], i.e., the assessment is indicated as gray-tone. As compared to the same set displayed in Figure 2c), color and eccentricity features are lost. These features are however used in the following search. Part b) of the figure shows the hundred best row-Gestalten that the greedy sequential search found here. Note, all of these are of much larger scale than their parts. The dominant almost diagonal structure is contained, and these are the longest rows found. Also, the shorter East-West row to the North of it is present in many variants. Part c) goes one step deeper in hierarchy. These are the reflection symmetries made from the rows. Note, such Gestalten are again larger in scale than their parts. In fact, they are about the largest possible objects with good assessment on such small image basis. A deeper hierarchy is only possible for reflections of reflections etc. But also with these the practical achievable depth is no deeper than five.

The drawing convention connects the parts of a reflection symmetry by the cross section line – the symmetry axis being perpendicular to that. The Gestalt content of image #5 is not very well reproduced by this set in Figure 3c). With rising depth of hierarchy often clusters result, while on the primitive level the objects are spread rather uniformly over the image. Accordingly, for the contests [11, 12] a cluster analysis completed the proposed procedure. The result is displayed in Figure 4.

The dominant cluster has a diagonal axis from left upper corner to lower right corner. This would not be consistent with the perceived Gestalt, the example #5 thus counting as failure. However, a cluster close to the perceived reflection pair of building rows is contained among the strongest clusters. It is slightly tilted in the other direction, from the horizontal direction, light grey, with its center about 550 units North and 400 units East of the origin.

Instead of using constant false alarm rate, one may also use a fixed limit on the assessments. Definitions such as (1) and (2) can be augmented by a bound stating, e.g. that the resulting Gestalten f must be better assessed than – say 0.5. If this way is pursued a certain degree of meaning will be ensured. Aggregates thus found in an image are of a certain minimal quality. However, the search efforts cannot be predicted anymore. Some images will give no Gestalt at all, and others, even if they are quite small, may contain huge numbers of possible hierarchical aggregates, where the listing becomes intractable.

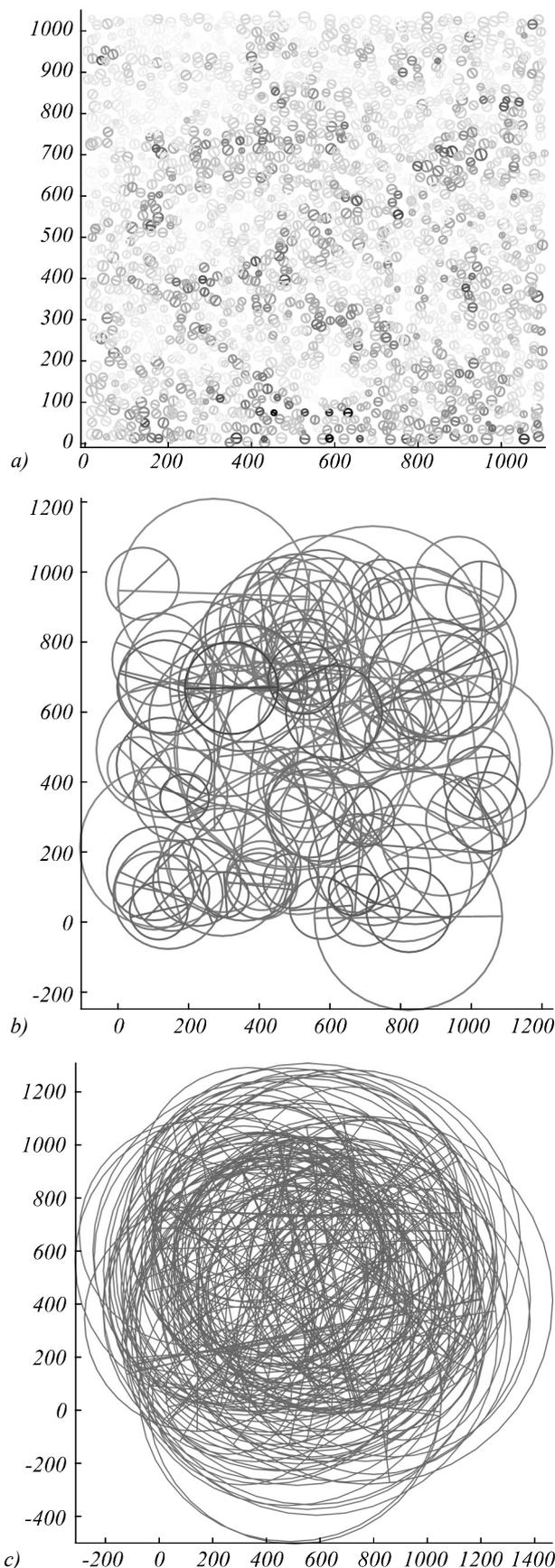


Fig. 3. Searching Gestalten on image #5 – a) primitives, b) $L_{100\Sigma}$, c) $L_{100|100\Sigma}$

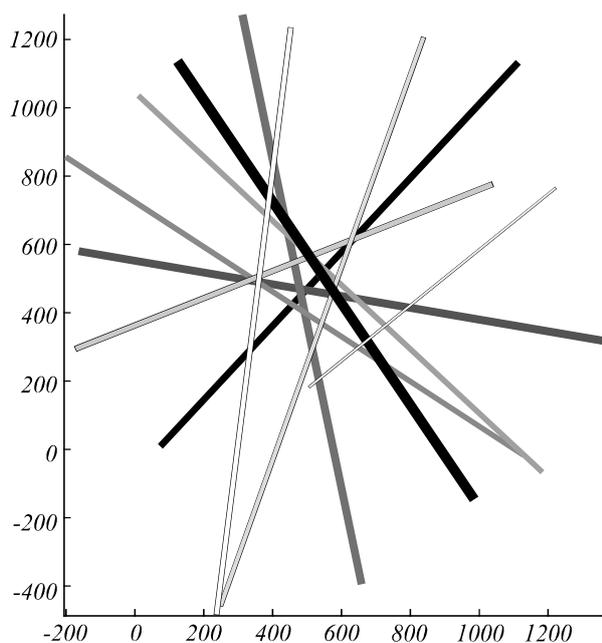


Fig. 4. Resulting clustered axes on image #5 – black: best cluster, grey: next best clusters, cluster-strength displayed as thickness and darkness

Top-down analysis

A lot can be learned from running such search on even only one image, such as #5 of the Thimphu set. It is essential to mark by a suitable graphical interface, what should be found in that image. Then the accumulated Gestalten found by the search are scanned for results similar to the ground truth. Figure 5 shows such a positive sample.

It is essential to record the parts with each aggregate, so as to understand the structure, and check consistency with the structure of the ground-truth. Both, ground-truth and objects found have a part-off-tree structure, the ground-truth tree having twelve leaves, and the object presented in Figure 5 having thirteen leaves. How can such trees be compared?

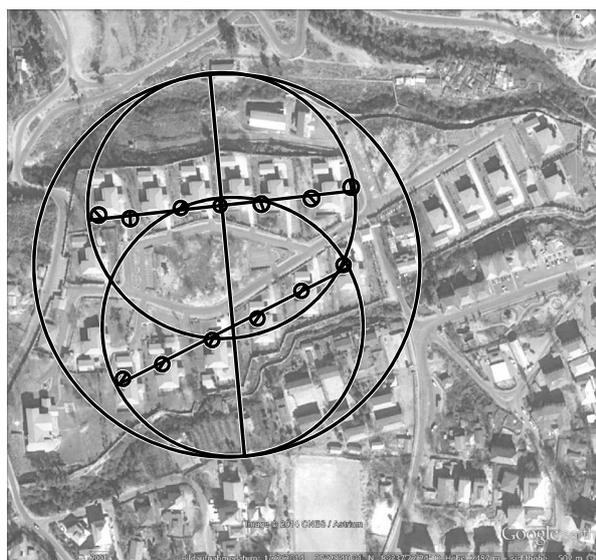


Fig. 5. Σ -Object closest to the ground-truth: a roughly reflection symmetric arrangement of two rows

For the time being, we should not be over-optimistic in order to avoid disappointment and frustration. Progress in this field comes in small steps. First of all, the assessment of the large reflection symmetry in Figure 5 is very bad. In this figure, we presented all Gestalten in black for better visibility over the brighter version of the super-pixel image, which is needed for reference. The two parts of this object are of similar size resulting in a good similarity-in-scale component of the assessment. However, they are too close to each other to be *in proximity*. The corresponding assessment function prefers objects that are in double distance of their scale. And there is a second bad component: The current search implementation also contains color propagation – the row-Gestalten inherit the mean-color of their parts, and *color-similarity* has currently a high weight in the assessment of reflection Gestalten. In the example at hand the Southern row is made from shadow-segments to the North of the buildings, which are dark, while the parts of the Northern row are light-grey segments between the buildings. The geometric assessment for reflection symmetry is also rather mediocre in this case; the orientations are not really mapped on each other.

Both row-Gestalten are among the two hundred best rows found. Color similarity and the *geometric good continuation assessment* are high. Also, the parts are spaced in close to optimal proximity. The only component lowering their overall assessment is *inheritance* of assessment from the parts. In particular, the parts forming the Northern row have low contrast to some of their neighbors and thus have low assessments.

All these problems can be mitigated by introducing or changing certain parameters, such as the preferred distance, different proximity assessments for reflections and rows, lower color similarity weight, etc. This can be done manually, provided there are suitable interactive interface tools for the operator. It can also be done automatically using learning rules, such as they are known from the artificial neural net community, or by assembling statistics, and parametrizing the assessment functions so that they fit to the densities found. Whether manual or automatic adjustment, the ground-truth must be provided by human subjects in any case. And in the end, it is preferred to use more than one image, in fact, a small but really representative set would be good.

There is also evidence for a more serious problem from the top-down analysis of the search on example Thimphu #5: The greedy row prolongation procedure only outputs the longest possible rows. It prolongs a row on both ends until the assessment would be significantly lower by the next best prolongation. Thus, e.g. the Northern row in Figure 5 has one element in the East too many. The semantically more suitable six-element row without this element might be an intermediate result, but will not be among the results stored. We follow here *A. Desolneux's* principle of the *maximal meaningful element* [2]. Note also, that all the rows that are made from roof segments along the oblique salient Southern building row formation are longer and contain additional building segments in good continuation further to the East. However, there is no alternative to the greedy search for the maximal meaningful element – listing all part-rows is not tractable.

Performance

Quantitative evaluation of our approach is a complicated issue. The usual publicly available remotely sensed benchmark data are labeled with land-use classes, such as vegetation, road surface, building, etc. They have no labels intended to capture *visual saliency*. Still, one possibility of Gestalt-recognition evaluation would be using one of the labeled classes, e.g., buildings, and evaluate the Gestalt algebra in cooperation with an object-recognizing method for this class. Then the gain in performance can be measured with Gestalt-grouping as compared to without such aid. Initial steps in that direction have been done in [19], where a self-organizing neural net was used on a hyperspectral benchmark in combination with Gestalt grouping.

Another possibility is in participating in symmetry recognition competitions, such as [11] and [12]. We did so, but with rather mediocre success. This does not mean that our methods are inferior. The data used for these evaluations were not representative for remote sensing. Faces are included as well as many animals, and some vehicles. There are some buildings in it, but captured from the ground, and often under strong perspective distortions. Furthermore, the categories of these competitions do not include hierarchical nested symmetries. *Liu* and her team stress the importance of this topic [11], but even defining a proper labeling scheme for such category would be challenging. Moreover, probably only one team could participate in it, because as far as we know, we are still alone with our approach.

Conclusion

Human observers are often superior in recognizing objects, and in particular in analyzing remotely sensed images of urban terrain. This is often attributed to their inherent common sense, which turns out one of the hardest problem of artificial intelligence research. However, with equal evidence, this remarkable performance might be due to Gestalt perception, and the laws of these latter capabilities might well be studied – also in application to remotely sensed data.

References

- [1] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S. SLIC Superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell* 2012; 34(11): 2274-2281. DOI: 10.1109/TPAMI.2012.120.
- [2] Desolneux A, Moisan L, Morel JM. *From Gestalt theory to image analysis: A probabilistic approach*. Berlin: Springer; 2008. ISBN: 978-0-387-72635-9.
- [3] Mumford D, Desolneux A. *Pattern theory: The stochastic analysis of real-world signals*. Natick, Massachusetts: CRC Press; 2010. ISBN: 978-1-56881-579-4.
- [4] Pizlo Z, Li Y, Sawada T, Steinman RM. *Making a machine that sees like us*. Oxford: Oxford University Press; 2014. ISBN: 978-0-19-992254-3.
- [5] Leyton M. *Symmetry, causality, mind*. Oxford: Oxford University Press; 1993. ISBN: 978-0-262-62131-1.
- [6] Gruen A, Kuebler O, Agouris P, eds. *Automatic extraction of man-made objects from aerial and space images*. Basel: Birkhaeuser; 1995. ISBN: 978-3-7643-5264-6.
- [7] Gruen A, Baltsavias EP, Henricson O, eds. *Automatic extraction of man-made objects from aerial and space images (II)*. Basel: Birkhaeuser; 1997. ISBN: 978-3-7643-5788-7.
- [8] Baltsavias EP, Gruen A, van Gool L, eds. *Automatic extraction of man-made objects from aerial and space images (III)*. Lisse: Balkema Publishers; 2001. ISBN: 90-5809-252-6.
- [9] *Image understanding workshop: Proceedings of a workshop held at Cambridge*. Two Volumes. Cambridge, Massachusetts: Morgan Kaufmann Pub; 1988. ISBN: 978-0-934613-68-2
- [10] Wertheimer M. Untersuchungen zur Lehre der Gestalt. II [In German]. *Psychologische Forschung* 1923; 4(1): 301-350. DOI: 10.1007/BF00410640.
- [11] Liu J, Slota G, Zheng G, Wu Z, Park M, Lee S, Rauschert I, Liu Y. Symmetry detection from real world images competition 2013: Summary and results. *The IEEE Conference on Computer Vision and Pattern Recognition Workshops 2013: 200-205*. DOI: 10.1109/CVPRW.2013.155.
- [12] Funk C, Lee S, Oswald MR, Tsokas S, Shen W, Cohen A, Dickinson S, Liu J. 2017 ICCV Challenge: Detecting symmetry in the wild. *2017 IEEE International Conference on Computer Vision Workshops (ICCVW) 2017: 1692-1701*. DOI: 10.1109/ICCVW.2017.198.
- [13] Stilla U, Michaelsen E. Semantic modelling of man-made objects by production nets. In: Gruen A, Baltsavias EP, Henricson O, eds. *Automatic extraction of man-made objects from aerial and space images (II)*. Basel: Birkhäuser, 1997: 43-52. DOI: 10.1007/978-3-0348-8906-3_5.
- [14] Michaelsen E, Jäger K, Roschkowski D, Doktorski L, Arens M. Object-oriented landmark recognition for UAV-navigation. *Pattern Recognit Image Anal* 2011; 21(2): 152-155. DOI: 10.1134/S1054661811020763.
- [15] Soergel U, Michaelsen E, Thiele A, Cadario E, Thoennessen U. Stereo analysis of high-resolution SAR images for building height estimation in cases of orthogonal aspect directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 2009; 64(5): 490-500. DOI: 10.1016/j.isprsjprs.2008.10.007.
- [16] Michaelsen E, Muench D, Arens M. Recognition of symmetry structure by use of Gestalt algebra. *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops 2013: 206-210*. DOI: 10.1109/CVPRW.2013.37.
- [17] Michaelsen E, Yashina VV. Simple gestalt algebra. *Pattern Recognition and Image Analysis* 2014; 24(4): 542-551. DOI: 10.1134/S1054661814040154.
- [18] Michaelsen E, Muench D, Arens M. Searching remotely sensed images for meaningful nested gestalten. *Int Arch Photogramm Remote Sens Spatial Inf Sci* 2016; XLI-B3: 899-903. DOI: 10.5194/isprsjprs-XLI-B3-899-2016.
- [19] Michaelsen E. Self-organizing maps and Gestalt organization as components of an advanced system for remotely sensed data: An example with thermal hyper-spectra. *Pattern Recogn Lett* 2016; 83(2): 169-177. DOI: 10.1016/j.patrec.2016.06.004.

Author's information

Eckart Michaelsen (b. 1958) graduated in mathematics from Innsbruck University in 1987, and obtained Dr.-eng. (PhD) in computer science from Erlangen University in 1998. Currently he works scientist in Fraunhofer-IOSB in Ettlingen, Germany. He has been chair of the IAPR-TC7 (Technical Committee for Pattern Recognition in Remote Sensing and Mapping) between 2014 and 2018. Research interests are knowledge-based image analysis, object recognition, remote sensing and perceptual grouping. E-mail: eckart.michaelsen@iosb.fraunhofer.de.

Received March 15, 2018. The final version – October 20, 2018.