Hierarchical Grouping Using Gestalt Assessments

Eckart Michaelsen, Michael Arens
Fraunhofer-IOSB, Gutleuthausstrasse 1, 76275 Ettlingen, Germany
eckart.michaelsen@iosb.fraunhofer.de

Abstract

Real images contain symmetric Gestalten with high probability. I.e. certain parts can be mapped on other certain parts by the usual Gestalt laws and are repeated there with high similarity. Moreover, such mapping comes in nested hierarchies – e.g. a reflection Gestalt that is made of repetition friezes, whose parts are again reflection symmetric compositions. This can be explicitly modelled by continuous assessment functions. Hard decisions on whether or not a law is fulfilled are avoided. Starting from primitive objects extracted from the input image successively aggregates are constructed: reflection pairs, rows, etc., forming a part-of-hierarchy and rising in scale. The work in this paper starts from super-pixel primitives, and the grouping ends when the Gestalten almost fill the whole image. Occasionally the results may not be in accordance with human perception. The parameters have not been adjusted specifically for the data at hand. Previous work only used the compulsory attributes location, scale, orientation and assessment for each object. A way to improve the recognition performance is utilizing additional features such as colors or eccentricity. Thus the recognition rates are a little better.

1. Introduction

In the symmetry recognition or Gestalt grouping community there is an ongoing dispute whether to use a set of certain primitive objects extracted from the image – like in [1], or to fill certain accumulators directly from the raw colors – like in Hough transform methods. The latter usually results in nested enumeration-loops, and may thus cause high computational efforts, while being conceptually fairly simple. The former will suffer from loss of information in the primitive extraction method.

If one decides for the primitive extraction way, there will be no reason to avoid nested hierarchies – except it may be a little bit more complex conceptually. With the term ‘nested hierarchies’ we refer to things like reflection symmetric arrangements of friezes of primitives, or rotation mandalas made up of reflection symmetric parts, etc.

Today pictures with ten or hundred mega-pixel are quite normal, and there are giga-pixel images around. The larger the image is, the more likely it is that it contains nested hierarchies of symmetries. On the other hand: To us it appears extremely unlikely that a real image – from the wild – may contain a region of white noise pixels in one larger region, and the same pixel configuration mapped by a symmetry law, such as reflection in a nearby region. Figure 1 shows an example from the benchmark data at hand. Although it is not very large, it contains several frieze symmetries on either side that are arranged in left-to-right reflection. Zooming in, we would furthermore realize that each window has again left-to-right reflection symmetry. The reflection symmetry dominates though actually there is one column of windows more on the right side.

Figure 1: Example #34 from the single reflection symmetry image set: Obviously there is a hierarchy of nested perceptual groups in this one

It is our intention to develop and test methods that may automatically find, parametrize, and assess such nested hierarchies. Obviously, the result should be some kind of parse-tree of a picture grammar. Given such structure was constructed, one may return to the raw colors of the pixels, comparing very specific small regions with each other, that may be arbitrarily far away from each other. This somehow reconciles the dispute between the primitive-extraction approach and the raw-data comparing.

1.1. Related work

Actually, these ideas are not very new, in particular for the analysis of remotely sensed data this was proposed already forty years ago [2], [3], when emphasis was still on knowledge-based approaches. Mathematical formulations for Gestalt laws, and their application in machine vision was best treated in [4], and the application of this theory to symmetry recognition competed with good success in the 2013 competition [5]. In the last decades, quantitative recognition success has been the focus of research attention. Somehow this draws attention away from the internal structure and organization of real images. Still there are exciting new findings on the topic by certain working teams. For instance, M. Irani's group found that in real images certain patches recur much more often, than can be expected from random images [6]. She emphasizes re-occurrence over different scales, and uses this property for foreground recognition, haze removal and so forth. Much work in field had façade analysis as focus [7] [8]. The most successful methods rely on grammars and sophisticated statistical sampling methods for the search. Façade recognition usually assumes ortho-rectified imagery, and prefers horizontal and vertical organization.

Our Gestalt hierarchic approach participated in the 2013 CVPR symmetry recognition competition, with rather limited success [9]. An algebraic foundation of such hierarchical Gestalt grouping in the form also used in this paper was attempted in [15]. This includes several theorems and lemmas that are also of practical relevance. Including SIFT 128-dimensional key-point features in order to improve the performance following [1] was demonstrated in [10]. Most of our papers using these methods concentrated on remote-sensing applications [11][12][13]. The images used were from very diverse sources such as SAR, hyper-spectral cameras, or satellite and aerial imagery. The clustering of assessed projective entities as outlined in Sect. 3.2.1 was first published for planes in 3D [19].

Section 2 of this paper introduces the important concepts of our approach, namely the Gestalt-domain, the operations on it, and the assessment functions coding the Gestalt-laws. For the application of this apparatus to specific visual recognition tasks two interfaces are needed: From the input image a set of small, primitive objects must be extracted that fit into the Gestalt domain, and from the set of accumulated larger aggregate Gestalten an output must be formed that fits the desired format. These interfaces are treated in Section 3. Very briefly Section 4 lists the experiments that were done for this 2017 competition, before Section 5 discusses the approach and gives an outlook on future work.

2. The Gestalt-domain

If certain image objects are to be arranged in nested hierarchies as outlined exemplarily above, it will be advisable to define a domain of compulsory features that such objects must have:

- All objects need a **location** in the image. For simplicity, the image margins will be ignored. Furthermore, no pixel raster is considered. Thus, we have the standard 2D-vectorspace as location domain, with all its benign algebraic and statistic properties.
- All objects need a **scale** (or size). Scales are positive, and they form a multiplicative group. We should never add or subtract scales. A scale should also never be assumed to be distributed normally (i.e., a Gaussian), because there cannot be negative scales. Instead one may consider e.g., Rayleigh distributions for scales. Also the mean of two scales should not be the arithmetic mean, it is more appropriate to use the geometric mean.
- All objects need an **orientation**. Algebraically, orientations are elements of an additive, continuous group. Gaussian distributions are not appropriate for orientation features. Instead e.g, Riess distributions can be used [16].
- All objects need an **assessment** between zero and one. Zero-assessed objects are meaningless and maximally-assessed are very salient.
- All objects need a rotational **frequency** (an integer >0). In this paper all examples have frequency *2*, because they are self-similar with respect to 180^0 rotations. But, e.g., if SIFT primitives are used, they will have frequency 1, and rotational Gestalten will be self-similar with rotations of $360^0/n$, *n* being the frequency.

Some objects may have **additional features**, such as colors, eccentricities, or arbitrary complex other properties, such as descriptor vectors commonly used in machine vision. Objects in this domain will be called **Gestalten**.

It is intended to test mutual geometric relations on configurations of such objects. The intuitive understanding of such relations is logical: Two or more orientations may be parallel or not, two or more scales may be equal or not. On the continuous domains at hand such exact equality is almost impossible (the probability for it is zero). So these relations will almost never be fulfilled, and almost no aggregate would possible. The standard way around this is the introduction of tolerance parameters: Two orientations are regarded as parallel if they do not differ more than – say 10^0.

Such ansatz leads to instable and unsatisfying behavior. Instead we prefer smooth **assessment functions** yielding a value between (and including) zero and one. One means perfect fulfillment of the relation, zero means perfect violation. This is similar to the well-known fuzzy-set membership functions. But we do not use piecewise linear ramps (with even more parameters). E.g. for parallelism in the 360^0 domain we prefer $½-½cos(\delta)$. One may also use functions of the form of Riess densities. For the time being

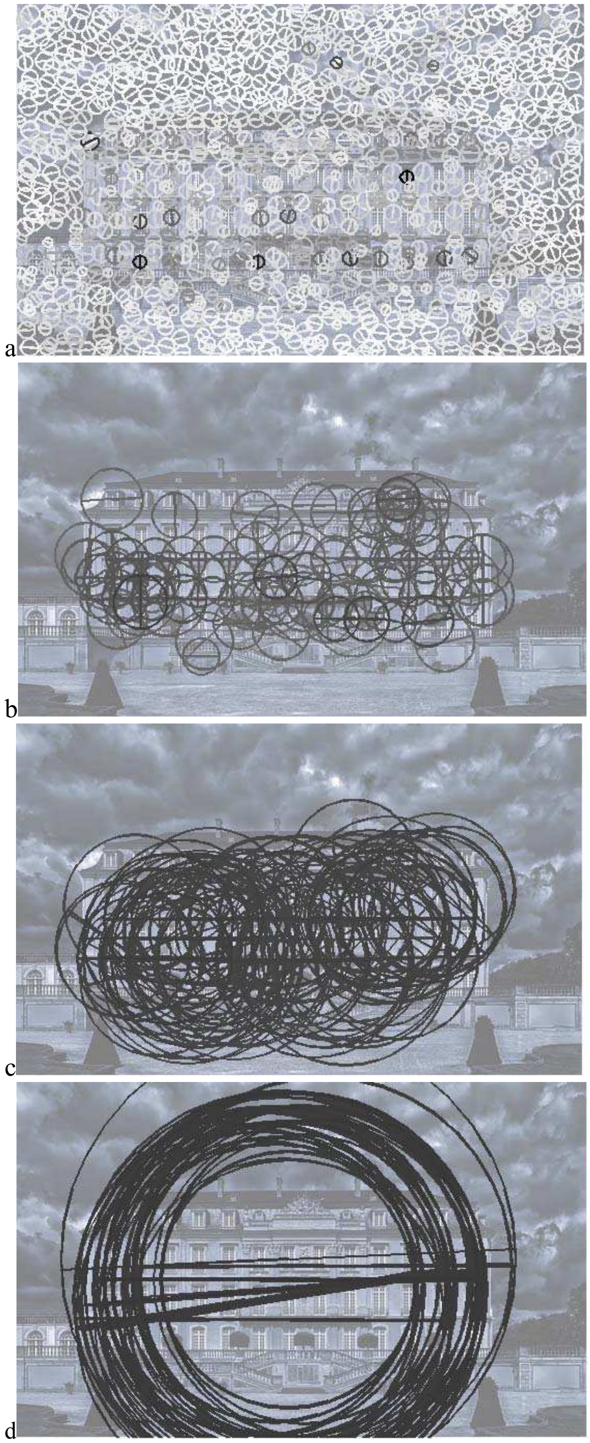


Figure 2: Applying the operation | successively on Gestalten extracted from an image: a: Primitives; b-d: level-1 to level-3 reflection Gestalten

we prefer differentiable functions with no parameters, but one may introduce parameters modifying the functions later in order to improve the recognition rates.

2.1. Reflection

A pair of Gestalten *(f,g)* forms a new aggregate Gestalt *h=f|g=g|f*, the **reflection** of *f* and *g*. It will be well assessed, if they are *close* to each other (i.e., in proximity), *similar* to each other in scale, and their orientation almost maps on one-another by the perpendicular bisector of the locations as *reflection axis*. Violating those Gestalt laws leads to a decline in assessment. Proposals for the corresponding continuous assessment functions were made in [9][15]. Since these functions range in *[0,1]* they may be considered as fuzzy memberships. Accordingly, multiplication of the assessments corresponding to the mentioned three laws (proximity, similarity, and reflection) to form a combined assessment is equivalent to a logical conjunction.

The Gestalten *f* and *g* also have assessments. So inheritance of assessment through the operation | is achieved by multiplying the outcome of the Gestalt-law assessments with the mid assessment of *f* and *g*. Additionally, if the Gestalten have additional features – such as colors, or eccentricity – similarity with respect to these may also be included in the overall assessment.

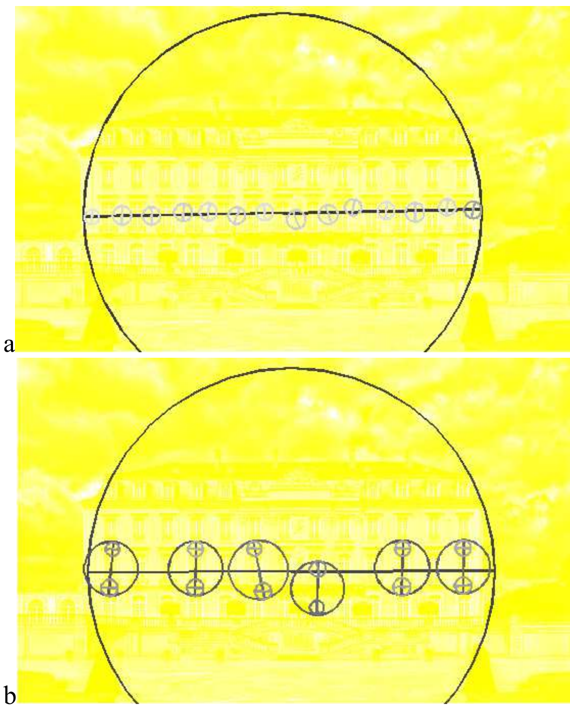
Thus, in the end, randomly picked Gestalten *f* and *g* will almost always cause a close-to-zero assessment of the aggregate *f|g*. It still exists, but it is quite meaningless. In this way, hard decisions following the Gestalt-laws based on thresholds are avoided. On the other hand, a set of Gestalten extracted from some given image, e.g. by some segmentation as outlined below will most often contain pairs that yield well assessed reflection Gestalten.

In Fig. 2 Gestalten are overlaid to the example image that were obtained by successive application of | to the primitives extracted from it. Drawing a Gestalt on an image is done according to the following convention: It is displayed as a circle with the circle-center at the *location* of the Gestalt, and the diameter is corresponding to its *scale*. All Gestalten in Fig. 2 are self-similar with respect to 180^{o} rotations. So the *orientation* attribute is drawn as diameter line. This line connects the two parts, so the symmetry axis would be perpendicular to this line. But it is not displayed. The *assessment* attribute is displayed as gray-tone – white meaning zero, and black meaning one.

Thus on a white background zero-assessed Gestalten disappear intentionally indicating that they are in fact meaningless. Displayed over a brightened version of the image they become visible again. Thus the reader can see almost all primitves on Fig.2a.

With rising level of hierarchy the number of possible Gestalten grows exponentially. If *n* is the number of Gestalten in one level – e.g. about 900 primitives in the example image – then there will be *n(n-1)/2* reflection Gestalten in the next level. Thus enumerating all level-4 Gestalten would cause considerable efforts. It can be demonstrated that with rising level only a very small portion of the possible Gestalten are assessed better than a given $\varepsilon>0$. Practically, it suffices to list all pairs of primitves, pick the hundred best of these level-1 |-Gestalten, form all pairs of these, pick again the best hundred from these level-2 |-Gestalten and so forth. One would also not accept anything worse than, say, 0.3.

In the example this process will terminate at the level-3, and the resulting set can be used for comparison with the ground-truth. If the image was bigger – and there was a similar second building in nice distance to the left or right of the building at hand, this symmetry would be found on level-4.



a

b

Figure 3: Examples of frieze-Gestalten on #34 of the single reflection data. a. Best level-1 frieze found – it is made of 14 primitives. b: Best level-2 frieze found – it is made of 6 |-Gestalten.

2.2. Frieze repetition

An n-tupel of Gestalten $(f_1,\ldots, f_n)$ forms a new aggregate Gestalt $g = \sum_{i=1\ldots n} f_i = \sum_{i=n\ldots 1} f_i$ the **frieze** or **row** of the f_i. It will be well assessed, if they are *close* to each other (i.e., in proximity), *similar* to each other in scale and orientation, and the locations are aligned in *good continuation*. Violating those Gestalt laws leads to a decline in assessment. Proposals for the corresponding continuous assessment functions were made in [9][15]. Again there is a logical conjunction of three laws (*proximity*, *similarity*, and *good continuation*) implemented as multiplication of the corresponding assessment functions, and again there is inheritance of the assessments form the parts to the aggregate.

Figure 3 exemplarily shows such objects, namely the best ∑-Gestalten found on hierarchy level one and two, respectively. As background we used an image with red- and green channel set to maximum, and only the blue channel taken from #34 (displayed in Fig.1). Thus in faint yellow the image structure is indicated while all Gestalten are clearly visible. In contrast to Fig. 2 we displayed only one (the best) aggregate Gestalt on each level, but we added the part-Gestalten from which it is constructed. Note that well assessed ∑-Gestalten grow faster in scale with rising level.

2.3. Rotational mandalas.

Configurations with rotationally symmetric arrangements of parts do not belong to the classical Gestalten as given e.g. in [17]. Yet it comes naturally to include them as an operation of its own: An n-tupel of Gestalten $(f_1,\ldots,f_n)$ forms a new aggregate Gestalt $g=\Pi_{i=1,2\ldots,n}\ f_i\ =\ \Pi_{i=n,1,\ldots(n-1)}\ f_i\ =\ldots=\ \Pi_{i=2,\ldots,n,1}\ f_i$ the **rotational mandala** of the f_i. It will be well assessed, if they are *close* to each other (i.e., in proximity), *similar* to each other in scale etc., and the locations and orientations are aligned in *rotational symmetry*.

Assessment functions, search procedures, and details about this operation have been given in [14]. For the time being, the search for these Gestalten causes considerable computational loads. Given the short time bounds, we cannot include them and participate in the rotational symmetry competition. But there are no conceptual obstacles prohibiting this, and we will process the rotational data given for the competition in due time.

2.4. Algebraic closure

Usually, in algebra, operations are considered that allow neutral elements or inverses. Not so for this case. It is easy to see that meaningful Gestalten (those with considerable assessment) will be larger in scale than their components. Also associativity is violated for these operations. They are no group operations at all. But the main reason why the term ‘algebra’ is used here, is *algebraic closure*: Any Gestalt can be combined with any other by any of the operations always yielding a new Gestalt inside the afore-mentioned domain. We gave a proof for that in [15]. So these operations are no productions of a grammar, where always only a subset of all combinations is admissible. Hierarchical grouping is done by forming a term not by parsing a word.

And more algebraic structure applies: We have seen above that *commutativity* holds for the Gestalt operations. For | it holds in its ordinary sense. For ∑ and ∏

commutativity needs to be generalized to *n*-ary operations. Here a sub-group of the permutations S_n is applicable without changing the content of the term, in the case of $\sum$ the two element group, and in the case of $\prod$ the finite rotational group. This can reduce the search effort: Only the combination corresponding to one member of the group needs to be evaluated.

Distributivity does not hold in general: E.g. $\sum g_i|f_i=\sum g_i|\sum f_i$ is true for location, scale, and orientation, but generally not for the assessment attribute. This may be fixed by defining that this is in fact one Gestalt and it gets the maximum of both assessments. If the orientation of a |-Gestalt is the same as the orientation of its parts – like in most cases in Figure 2 – or perpendicular to it, it will also be a $\sum$-Gestalt with two members, and if we have nested |-applications in such cases it will be a $\sum$-Gestalt with 2^k members (where k is the hierarchic depth). Most $\prod$-Gestalten that occur on real data have in fact di-eder symmetry, so there is a whole group of |-terms for them as well. We are just starting to investigate this algebraic structure and its implications on the search complexity.

2.5. Transporting evidence through a nested term

When looking at a nested terms such as the $\sum g_i|f_i$ Gestalt in Figure 3b one can imagine that an interrelation between far away primitive objects – such as the leftmost lower primitive g_1 and the rightmost upper element f_6 is constructed, something like a correspondence. E.g. if they are of similar color and eccentricity there is more evidence for the validity of the whole aggregate – and vice versa. Indeed such transport of additional attributes through the hierarchy is now newly included, and probably helps improving performance.

On the other hand, orientations are not propagated yet. Note, that the orientation feature is compulsory in the Gestalt domain. The orientation of the aggregate is horizontal in the example. If we included the mid-orientation of its parts (in this case vertical) as additional feature, and the orientation of their parts again, we would end up with the number of additional features increasing with increasing hierarchical depth. This is not implemented yet, but properly done, such additional comparisons should punish the inclusion of parts such as $g_4|f_4$ in the example, where the orientation is vertical, instead of horizontal like with all other primitives.

However, when including more and more punishment, care has to be taken, that at least some non-trivial Gestalten survive. Artificial illusion has to be considered for cases, where a structure in otherwise perfect continuation features a gap.

3. Incorporating Gestalt search into a solution

For almost all applications, a set of objects in the Gestalt domain is neither given as input datum, nor is such set a proper output. For the symmetry recognition competition at hand the input format is a (mostly colored) picture given in a pixel grid, and the output format are a few locations, e.g., for single reflection the begin- and end-location of the estimated axis. These are compared with the ground-truth that comes with the data.

3.1. From the picture to the primitive Gestalten

Most work on hierarchical Gestalt grouping used the well-known SIFT key-points as primitives [9][10][11][12]. The reason might have been, that SIFT points provide exactly the desired Gestalt domain features *location*, *scale*, *orientation*, and *assessment*, and maybe also because the standard solution of Loy & Eklundh [1] also was based on these.

However, such SIFT key-points rarely correspond to objects in the image in accordance with human perception. They rather appear at corners of such objects or somewhere in textured regions. We therefore looked for a method yielding primitives that are more in accordance with human segmentation, and found the SLIC super-pixel segmentation method [18].

Figure 4a shows the result of segmenting super-pixels from #34 of the single reflection data at hand for this competition. For each super-pixel the Gestalt domain features *location*, and *scale* are straightforward. The *orientation* is set from the second moment of the object. It may be instable, if the object should turn out isotrop.

For a Gestalt also an *assessment* is required. Note, that super-pixels surrounded by neighbors with the same or similar colors are meaningless in their location, scale, etc. They just reproduce the hexagonal grid. Accordingly, we set the assessment for such object to zero. A super-pixel with maximal color difference to its neighbors will be assigned with assessment one, and in between some continuous function is used. In the figure, most meaningless super-pixels have very dark colors and thus disappear on the black ground on which the objects are displayed.

One very important advantage of such extraction method is that such intuitive display is possible, so as to estimate how much information is lost, and rate the accordance with human perception. Like most primitive extraction methods, super-pixel segmentation comes with parameters, and the overall success will depend on a suitable choice of these. With a set of images and corresponding ground-truth at hand, one may also optimize the performance by varying these parameters – provided there was enough time.

3.2. From the accumulated Gestalt-set to the output

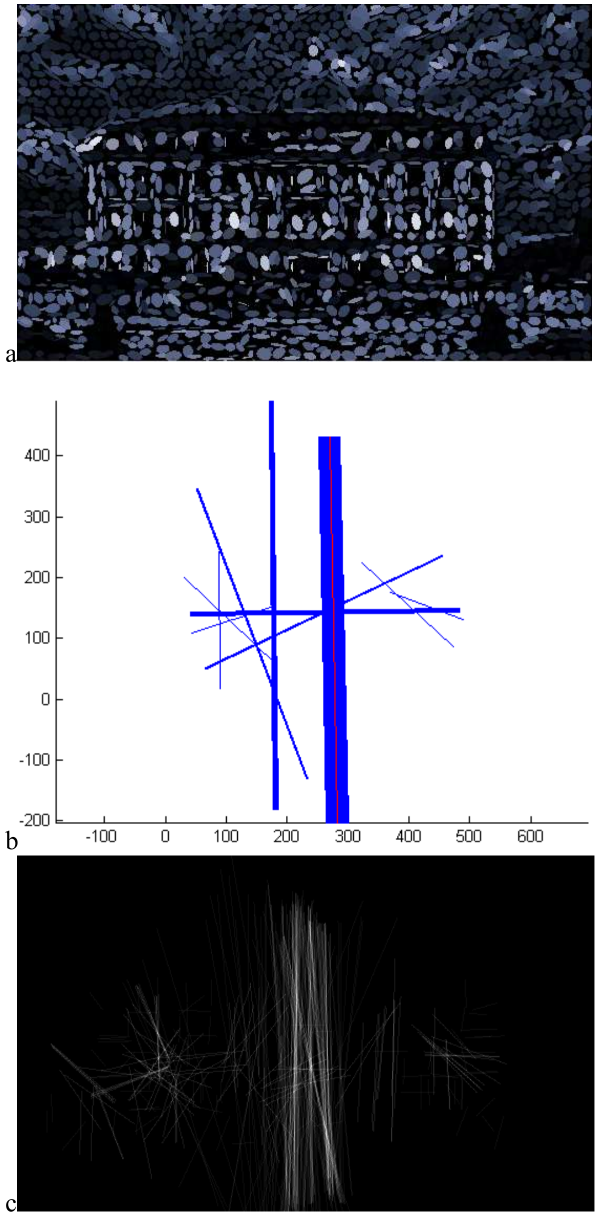


Figure 4: a. SLIC-segmentation of #34 refl. sing.: Only location, size, orientation, color, and eccentricity remain for each primitive. Note: There is not much color in this one. b. Axis clustering for the reflection symmetry contest. Decision on #34 is for the dominant central vertical element which is marked by a thin red axis accordingly. c. Reflection symmetry heatmap on #34 according to the CoCo specification.

Simply picking the best assessed Gestalt, and transform its features to the output format required for the competition, does not perform well. Therefore, a cluster procedure is used, that regards |-Gestalten, whose axes are roughly collinear, as mutually affirming. This comes fairly close to the usual Hough-accumulation used for many state-of-the-art approaches to reflection symmetry recognition. Of course, different methods are used for the different competitions:

3.2.1 Single reflection

The proper domain for reflection axes in 2D is the projective plane $\mathbf{RP}^2$ - neither a vector-space nor a metric space – and the elements of it are written as homogenous triples $a=(a_1,a_2,a_3)$. A kind-of-distance between two elements of this domain a and a' can be found by scaling the coordinates of both so that $a_1^2+a_2^2=1$ (which is not possible for the line-at-infinity, not occurring in our application) and then taking the Euclidean vector distance between these coordinates or the distance with one sign flipped:

$$d(a,a')=min(||a-a'||,||a+a'||)$$

It is known that this pseudo-distance seriously varies with the choice of the coordinate system. We follow here [20] transforming the coordinates such that the origin is the image center, and the smaller image dimension sets length 1. Using this setting, a certain weighting between deviations in orientation and deviations in set-off is chosen. The distance between two objects will be different if they are close to the image-center as compared to somewhere close to the margin. We are well aware, that these constructions violate our ansatz to treat every Gestalt equal, no matter where in the image it appears.

By means of a suitable threshold τ we may form a cluster of mutually consistent axes in a set of given axes. Since our Gestalten are additionally attributed by an assessment, we may start with the best assessed, and proceed as follows:

1. Pick the best a_i and count all a_j with $d(a_i,a_j)<\tau$. This count will serve as accumulated evidence for the corresponding cluster.
2. Re-assess all axes using a monotone function of $d(a_{i,},a_j)$, which yields zero-out for zero-in and one-out for maximal possible inputs. Thus, for instance a_i will be assessed zero, and now axes perpendicular to it, or with a very different offset will rise in the assessment rank-order.
3. Continue with step 1, either for a fixed number of steps (say ten), or the assessments sink below a threshold.

Figure 3b shows that often such clustering results in one very dominant cluster, in this case (#34) a vertical reflection axis through the center of the image. The thickness of the lines represents their accumulation value. The best element – that serves as output is additionally marked in red color. Note, that end- and begin-locations along the axis are chosen according to the size of the underlying Gestalten, which are imagined circular. Thus, the best axis cluster turns out to be longer than the ground-truth given for the contest regularly. Since this spoils the quantitative recognition performance, we decided to shorten all outputs by factor 0.6. For this example, this may suffice to be counted as success.

Figure 3c gives the heat-map corresponding to #34 in the specification given for the CoCo contest. For this no clustering in the projective domain of axes is necessary. Therefore we would prefer this kind of interface. It is also very intuitive to the human view, and loses no information apart from that induced by the pixel raster. Also it accepts no evidence outside the image margins. The step from a |-Gestalt to a CoCo-entry is again a rotation by 90 degree (since the axis is meant and not the connecting line) and a down scale again by the global factor 0.6. Such an entry than gives a line with begin- and end- location. Along this line the pixel-pins are incremented. Finally, in order to meet the norming, the picture is divided by its maximal entry.

3.2.2 *Multiple reflection*

The axes clustering outlined above outputs a set of axes-clusters. The first ten entries of this list for #34 of the single reflection data are displayed in Fig. 3b in blue color with thickness indicating accumulated evidence (the best in red). Decision for a set of output elements, as demanded for the multiple reflection contest, can be controlled by a minimal ratio between the best and the accepted, or by an absolute threshold for the accumulation. In the example #34 (single reflection data) any reasonable choice would probably decide for a one-element output.

3.2.3 *Frieze repetition*

For friezes also an output-clustering is required. However, frieze clustering is clustering in a vector space. First of all, only ∑-Gestalten with the same number of parts *n* will be clustered. Then both, the *location*, as well as the *generator* vector, that shifts the location of one part to the location of the next part, should be similar. This is simple clustering in 4D vector-space, which is a metric space. The resulting clusters have the same attributes, and these can be used to construct a *2*n* grid point raster conform with the ground-truth format.

3.2.4 *Heat-map*

The CoCo part of the 2017 ICCV competition uses a raster-map of the 400x400 entries between zero and one as ground-truth. Looking at Fig. 2 the reader may guess that such format is rather straight-forward for hierarchical Gestalt operation search. For instance for rotational symmetry, one may just round the location of the best – say 500 – accumulated Π-Gestalten, and increment a zero initialized field at these cells. Then the result must be normed by dividing through the maximum.

For the reflection part the locations alone do not suffice. Again we have to add 90 degree to the orientation of the best accumulated |-Gestalten, so that a line segment is constructed that visualizes the symmetry axis. The length of this line is again shortened by factor 0.6. Along this line locations are enumerated in one pixel distances, rounded, and the corresponding cells are finally incremented. In the end the result is normed again.

4. Experiments

The following data were used: Single Reflection (100 images), Multiple Reflection (100 images), and CoCo-Reflection (250 images). Primitive extraction and hierarchical Gestalt search was the same for all experiments. The CoCo-data feature strange white bars either at the top and bottom, or at the left and right margins. The hierarchical Gestalt accumulation would ignore super-pixels in the interior of these areas, but most often there is a strong edge, where the real image starts. And this 400 Pixel long strong straight contour will almost always dominate the result. Therefore, a small function was created that removes these stripes. The smaller format result is later pasted into the 400x400 map at the corresponding location.

We do not claim any quantitative recognition rates here, and leave the evaluation of the experiments to the competition team. We can also provide our code, so that our function can be evaluated on other data.

5. Discussion

Reproducing the hierarchical nested Gestalt perception of the human vision system turns out a hard challenge. Often the machine solution behaves counter-intuitive. Creating substantial ground-truth in the format in which our hierarchical Gestalt is depicted in Fig. 3b, would require very many clicks.

We emphasized, that the assessment functions should be continuous and differentiable. They can be parametrized. We used, e.g. a parameter for the law of proximity (similar to the parameter of a Rayleigh distribution). This was set by hand to 2.0 using sparse testing on one or the other image. We can imagine methods to adjust such parameters by gradient ascent utilizing partial derivatives of the empirical success with respect to such parameters. This would be very similar to back-propagation learning, the difference being that the hierarchical Gestalt grouping will only have four or five such parameters, so that a set of a hundred test images with (possibly hierarchical) ground-truth could result in very stable and optimal parameter settings. At least on this learning set then the performance would be much better – and probably also on unseen imagery.

For the time being, it is our experience that the performance of the hierarchical Gestalt grouping abused as simple symmetry detector is sufficient on imagery with no projective foreshortening, and other interfering circumstances, such as lighting, or partial occlusion. #34 of the single reflection data is a good example. Projective foreshortening can be included explicitly into the law of good continuation if desired.

It is also our experience that the system fails to reproduce the CoCo ground-truth on almost all corresponding images. Often the axes found by the search for nested |-Gestalten are rather perpendicular to the CoCo ground-truth. Yet, we do not want to disclose this negative result from publication. It may yield a bad place in the rank-order of the competition. Still, it can be of interest for the community.

Many of the images in the CoCo set contain people or objects of interest for people. Thus person-detectors and object-recognition methods trained on such objects may perform well on this set – without using any concept of symmetry.

References

[1] G. Loy and J.-O. Eklundh: Detecting symmetry and symmetric constellations of features. In Proc. of ECCV, 2006.

[2] T. Matsuyama and V.-S. Hwang: Sigma - A Knowledge-based Image Understanding System, Plenum Press, New York, 1990.

[3] M. Nagao, T. Matsuyama: A Structural Analysis of Complex Aerial Photographs, Plenum Press, NewYork, 1980.

[4] A. Desolneux, L. Moisan, J.-M. Morel: From Gestalt Theory to Image Analysis, Springer, New York, 2008.

[5] V. Patraucean, R. Grompone von Gioi: Detection of Mirror–Symmetric Image Patches. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR 2013, Portland, Oregon, Vol.1: 211-216.

[6] M. Zontak, I. Mosseri and M. Irani: Separating Signal from Noise using Patch Recurrence Across Scales, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2013.

[7] S. Wenzel: High-Level Facade Image Interpretation using Marked Point Processes. PhD thesis, Univ. of Bonn, Germany, 2016.

[8] R. Tylecek. Probabilistic Models for Symmetric Object Detection in Images. PhD thesis, Czech Technical Univ. Prague, 2016.

[9] E. Michaelsen, D. Muench, U. Arens: Recognition of Symmetry Structure by Use of Gestalt Algebra. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR 2013, Portland, Oregon, Vol.1: 206-210.

[10] E. Michaelsen. Gestalt Algebra - a Proposal for the Formalization of Gestalt Perception and Rendering. Symmetry, Vol. 6, No.3, pp. 566-577.

[11] E. Michaelsen, R. Gabler, N. Scherer-Negenborn. Towards Understanding Urban Patterns and Structures. PIA & HRIGI joint ISPRS Conferences, Munich, March 2015, Archives of ISPRS, http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-3-W2/135/2015/isprsarchives-XL-3-W2-135-2015.html.

[12] E. Michaelsen, D. Münch, M. Arens. Searching Remotely Sensed Images for Meaningful Nested Gestalten. XXIII ISPRS Congress 2016. Commission III, Praque, 2016, pp. 899-903

[13] E. Michaelsen. Self-organizing maps and Gestalt organization as components of an advanced system for remotely sensed data: An example with thermal hyper-spectra. Pattern recognition letters 83 (2016), Pt.2, pp.169-177.

[14] E. Michaelsen. Searching for Rotational Symmetries Based on the Gestalt Algebra. 9-th Open German-Russian Workshop on Pattern Recognition and Image Understanding, Koblenz, December 2014.

[15] E. Michaelsen, V. V. Yashina: Simple Gestalt Algebra. Pattern Recognition and Image Analysis, 24 (4), 2014: 542-551.

[16] N. I. Fisher: Statistical Analysis of Circular Data. Cambridge Univ. Press, Cambridge, New York, 1993.

[17] M. Wertheimer: Untersuchungen zur Lehre der Gestalt, II. Psychologische Forschung, 4, 301-350, 1923.

[18] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Susstrunk: SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. PAMI, 34 (11): 2274-2281.

[19] W. von Hansen, E. Michaelsen, U. Thoennessen: Cluster Analysis and Priority Sorting in Huge Point Clouds for Building Recognition. ICPR 2006, Hong Kong.

[20] R. Hartley, A. Zisserman, A. Multiple View Geometry. Cambridge University Press, Cambridge, 2000.