Hierarchical Grouping - the Gestalt Assessments Method

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

Abstract

Real images contain reflection symmetry and repetition in rows 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. It is our intention to develop and test methods that may automatically find, parametrize, and assess such nested hierarchies. This can be explicitly modelled by continuous assessment functions. The recognition performance is raised utilizing additional features such as colors. This paper reports examples from the 2017 data set.

1. Introduction

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. But even in images of moderate size they may be found. Figure 1 shows an example from the benchmark data at hand for the 2017 symmetry competition along with the ICCV (namely #34 of the single reflection data). 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 left side.

An automatic procedure for the recognition of such Gestalt structure should give some kind of parse-tree of a picture grammar. This short paper explains the method briefly. It is in no way optimized yet to yield high recognition rates in the competitions. E.g., there are several parameters in the method, which have been chosen rather preliminarily and arbitrarily. There proper adjustment will be a topic of future work.

1.1. Related work

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]. 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.

Figure 1: An example image (#34 single reflection)

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 papers using these methods concentrated on remote-sensing applications [11][12][13]. The clustering of assessed projective entities as outlined in Sect. 3.2.1 was first published for planes in 3D [19]. The greedy search in 2.2 was described in more detail in [12].

2. The Gestalt-domain and some operations on it

We use the following domain: All objects need a **location** in the image. All objects need a **scale** (or size). Scales are positive, and they form a multiplicative group. All objects need an **orientation**. Algebraically, orientations are elements of an additive, continuous group. All objects need an **assessment** between zero and one.

Zero-assessed objects are meaningless and maximally-assessed are very salient.Some objects may have **additional features**, such as colors, eccentricities, or arbitrary complex other properties.

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]. We used $\lambda(\alpha d)\cdot exp(-(\alpha d)^2)$ as *proximity assessment*. λ is a norming constant so that the maximal assessment is one. α is a parameter chosen 2 for these experiments, and d is the Euclidean distance between the locations. This proximity assessment has the form of a Rayleigh density. For *similarity in scale* we used $exp(2-s_f/s_g-s_g/s_f)$ which turns out one for equal scales, and tends to zero if the scales s_g and s_f are more different. There is no parameter in this assessment. The natural choice for *reflectivity* assessment is $½-½cos(o_g+o_f -2o_{g|f})$ if self-similarity with respect to 180^0 rotation is given and o refers to the orientation between 0^0 and 180^0.

Inheritance of assessments through the operation | is achieved by multiplying the outcome of the Gestalt-law assessments with the geometric mean 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.

In Fig. 2 Gestalten are overlaid to the example image that were obtained by successive application of **|** to the primitives extracted from it. Drawing uses the following convention: A circle is displayed with the center at the *location* of the Gestalt, the diameter is corresponding to its *scale*, *orientation* attribute is drawn as diameter line (we have self-similarity with respect to 180^0 rotations), *assessment* is displayed as gray-tone – white meaning zero, and black meaning one.

Practically, it suffices to list all pairs of primitives, 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. Here this will terminate at the level-3.

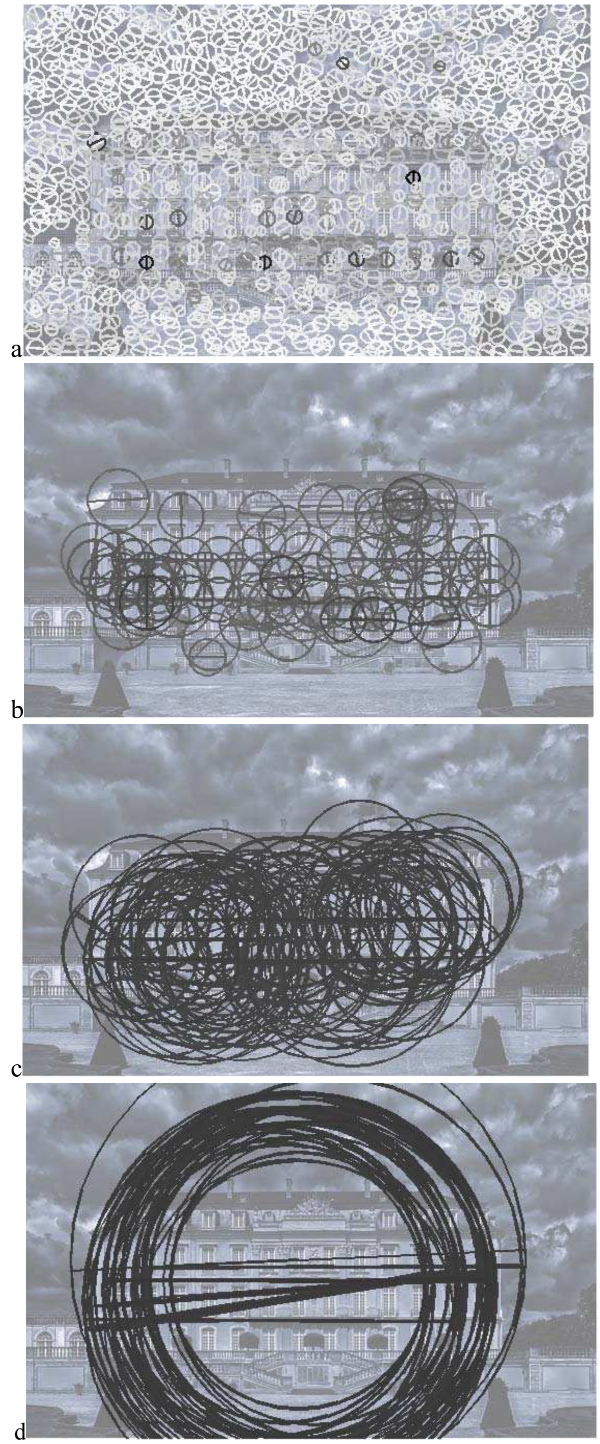


Figure 2: Applying the operation | successively on Gestalten extracted from #34: a: Primitives; b-d: level-1 to level-3 reflection Gestalten

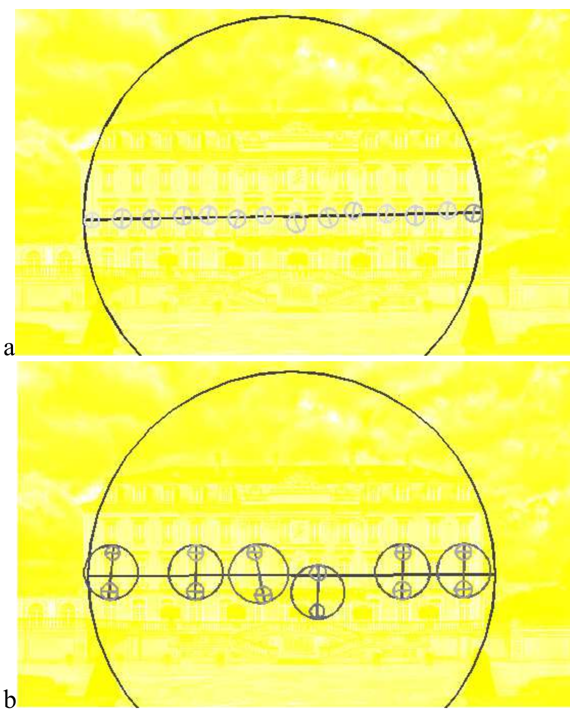
2.2. Frieze repetition

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*. Again there is multiplicative conjunction of three laws (*proximity*, *similarity*, and *good continuation*) and again there is inheritance of the assessments form the parts to the aggregate. The *good continuation* assessment is obtained as $exp(-\Sigma_{i=1\ldots n}\ \beta(\delta_i/s_i)^2\ )$, where the δ_i are residuals between the set-position of an optimal row and the real locations. β is again a parameter chosen as *2* for the time being.

While listing all pairs on each level for the |-operation is feasible, listing all *n*-tupels on each hierarchy level for the Σ-operation is not feasible. This would mean enumerating the power-set at each level. Instead a greedy search is performed: First all pairs Σg_1g_2 are tested. Then all which are better than a threshold are greedily prolonged at the end – we choose that g_3 that yields the best $\Sigma g_1g_2g_3$. The same is done at the beginning: choose that g_4 that yields the best $\Sigma g_4g_1g_2g_3$. This is repeated until assessments are getting worse. In the end multiple listings of the same *Σ*-Gestalten need to be removed. There is no guaranty that this procedure finds the optimum, but it is a good heuristic.

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. 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.

3. Incorporating Gestalt search into a solution



a

b

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

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 specific output formats are required.

3.1. From the picture to the primitive Gestalten

We 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.

3.2. From the accumulated Gestalt-set to the output

Cluster procedures are used, that regard e.g. |-Gestalten, whose axes are roughly collinear, as mutually affirming. 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.

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:

a) 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.

b) 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.
c) Continue with step a), 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.

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. 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.

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. 4b 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.

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 should be similar. This is simple clustering in 4D vector-space. The result is used to construct a $2*(n+1)$ 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 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.

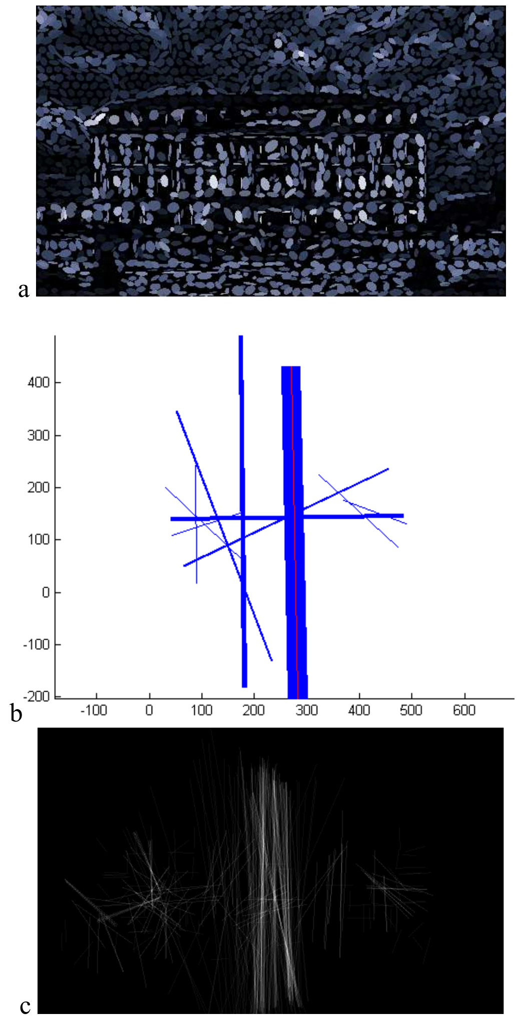


Figure 4: a. SLIC-segmentation of #34, b. Axis clustering for the reflection symmetry contest. The best is marked by a thin red axis accordingly. c. Reflection symmetry heatmap on #34 according to the COCO specification.

The COCO -data feature strange white bars either at the top and bottom, or at the left and right margins. 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.

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, M. 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.