Recognition of Symmetry Structure by Use of Gestalt Algebra

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

Abstract

While most approaches to symmetry detection in machine vision try to explain the gray-values or colors of the pixels, Gestalt algebra has no room for such measurement data. The entities (i.e. Gestalten) are only defined with respect to each other. They form a generic hierarchy, and live in a continuous domain without any pixel raster. There is also no constraint forcing them to completely fill an image, or prohibiting overlap. Yet, when used as a tool for symmetry recognition, the algebra must be somehow connected to the given data. In this paper this is done only on the primitive level using the well-known SIFT feature detector. From a set of such SIFT-based Gestalten follows a combinatorial set of higher-order symmetric Gestalten by constructing all possible terms using the operations of the algebra. The Gestalt domain contains a quality or assessment dimension. Taking the best Gestalten with respect to this attribute and clustering them yields the output for this competition participation.

1. Introduction

In man-made objects as well as in natural scenes symmetry comes in hierarchies. E.g. in an aerial picture of an urban scene the large scale symmetry might be a rotational mandala, while the parts (leafs) of this pattern appear on a smaller scale as frieze symmetry (a street with rows of equally shaped buildings on both sides), each building may again on a smaller scale be mirror symmetric, etc. Describing such a scene only on one of the scales falls too short.

1.1. Related work

Liu et al. [6] emphasize the algebraic nature of symmetry. Yet, up to now, most methods for symmetry recognition in machinevision concentrate on simple straight-forward accumulators implementing but one possible relation, and concentrating on recognition rates [2]. In the photogrammetric community some syntactic approaches have been proposed for urban data [5, 8, 11]. Still, generic models that can reduce descriptions in an infinite language by a finite set of rules, are rare in machine vision.

This is in contrast to the symmetry rendering community in computer graphics. Here many syntactic approaches have been published capturing more sophisticated algebraic structure and generic depth in their production rules, e.g. [4].

Next to generic hierarchy our ansatz emphasizes the continuous nature of deviation from the ideal setting of pattern symmetry, possibly with only one quality measure. This is common with the work on *near regular textures* (NRT) [6].

Of particular interest are assessment functions weighting the mutual position of parts of an aggregate, as given in this work by equations (3) and (8). They should not be too close and not too far from each other (also depending on their scale). This is motivated by successful computer vision grouping methods mimicking visual attention with local excitation and inhibition. An example of such work citing according psychological experiments is [1].

1.2. Previous own work

The Gestalt algebra has been defined in [9] giving closure theorems and useful lemmas. Yet, that work did not present any empirical results. In a way it is a consequence from experiences with structural or syntactic symmetry recognition on various remote sensing data [8].

2. Gestalt algebra and how to use it on images

The gestalt domain is defined as a manifold with margins. Beside the image position it contains scale, orientation, and assessment. We only give a short overview and refer to [9] for proofs of closure, related work, and additional lemmas. Then, we sketch how the primitive Gestalten extracted from an image can be combined in algebraic terms and listed. The best list entries are the output.

2.1. Gestalt algebra

As Gestalt domain we set

$$G = \mathbb{R}^2 \times (\mathbb{R} \bmod \pi) \times (0, \infty) \times [0,1], \qquad (1)$$

containing for a Gestalt $g \epsilon G$ position $po(g) \epsilon R^2$; orientation $or(g) \epsilon R \bmod \pi$; scale $sc(g) > 0$; and assessment $0 \leq as(g) \leq 1$.

Two operations are defined for the algebra up to now, and the corresponding proofs of algebraic closure are given in [9]. The first operation is binary, and written as |. It is meant for **mirror symmetry** and defined as

$$g \mid h = \begin{pmatrix} 1/2\left(po(g)+po(h)\right) \\ ori\left(po(g)-po(h)\right) \\ \mid po(g)-po(h) \mid +\sqrt{sc(g)sc(h)} \\ \left(a_{\mid,p} \cdot a_{\mid,o} \cdot a_{\mid,s} \cdot a_{\mid,a}\right)^{\frac{1}{4}} \end{pmatrix}, \qquad (2)$$

where $ori(v)=atan(v_y/v_x)$ and the assessment is calculated using (3) to (6):

$$a_{\mid,p} = e^{2-\frac{|po(g)-po(h)|}{\sqrt{sc(g)sc(h)}}-\frac{\sqrt{sc(g)sc(h)}}{|po(g)-po(h)|}} \qquad (3)$$

$$a_{\mid,o} = \frac{1}{2}+\frac{1}{2}\cdot\cos\left(2\left(or(g)+or(h)-2or(g\mid h)\right)\right) \qquad (4)$$

$$a_{\mid,s} = e^{2-sc(h)/sc(g)-sc(g)/sc(h)} \qquad (5)$$

$$a_{\mid,a} = \sqrt{as(g)as(h)}. \qquad (6)$$

Through (3) inhibition and excitation are combined preferring partners that are in good mutual distance. This is motivated by the success of visual machine-attention as e.g. published in [1]. Only through (4) mirror symmetric settings according to Wertheimer's laws [12] are preferred. Actually capturing mirror symmetry on orientations in this way has been proposed for computer vision e.g. in [10].

The second operation $\sum$ is of arity $n>1$. So n Gestalten are parts of a larger aggregate here. It is meant for **good continuation in rows** and defined as

$$\sum \boldsymbol{g} = \sum g_1 \ldots g_n = \begin{pmatrix} 1/n \cdot \sum po(g_i) \\ ori\left(\sum\left(po(g_{i+1})-po(g_i)\right)\right) \\ sc_{mid}+\mid po(g_n)-po(g_1)\mid \\ \left(a_{\Sigma,p}\cdot a_{\Sigma,o}\cdot a_{\Sigma,s}\cdot a_{\Sigma,a}\right)^{\frac{1}{4}} \end{pmatrix} \qquad (7)$$

with sc_{mid} as the geometric average scale and the assessment is calculated using (8) to (12):

$$a_{\Sigma,p} = \left[\prod_{i=1}^{n} e^{-\frac{|po(g_i)-set_i|}{sc(g_i)}}\right]^{\frac{1}{n}} \qquad (8)$$

$$set_i = po\left(\Sigma\boldsymbol{g}\right)+\left(-\frac{1}{2}+\frac{i-1}{n-1}\right)\left(po(g_n)-po(g_1)\right) \qquad (9)$$

$$a_{\Sigma,o} = \left(\prod_{i=1}^{n}\left[\frac{1}{2}+\frac{1}{2}\cos\left(2\left(o(g_i)-avo(g_1\ldots g_n)\right)\right)\right]\right)^{\frac{1}{n}} \qquad (10)$$

$$a_{\Sigma,s} = e^{(2n-t_1-\ldots-t_n-1/t_1-\ldots-1/t_n)/n}, \qquad (11)$$

where, similar to (5), t_i is the ratio between $sc(g_i)$, and the generator length $1/n\cdot|po(g_n)-po(g_1)|$.

$$a_{\Sigma,a} = \left(as(g_1)\cdot\ldots\cdot as(g_n)\right)^{\frac{1}{n}} \qquad (12)$$

Only through (8) and (9) good continuation into row settings according to Wertheimer's laws [12] are preferred.

2.2. Searching for Gestalten

We start by extracting primitives from the input images using the SIFT feature detector. There is a critical threshold controlling the number of primitives and what information we lose. SIFT fits well to our Gestalt domain (1). Position and scale are determined. For the scale we need a global factor describing how big a primitive SIFT Gestalt is in relation to the scale level it is found in. We found that a global factor of around 5 is plausible. The orientation is obtained from the maximum value in the gradient orientation histogram. Following this we construct successively Gestalt algebra terms using the following recursive algorithm:

1. Insert all primitives to a list L_0. Select $\varepsilon>0$.
2. Form terms $g\mid h$ from $\{g,h\}\epsilon L_0$ and store those with $as(g\mid h)>\varepsilon$ in a list L_{m1}.
3. Form terms $\sum gh$ from $\{g,h\}\epsilon L_0$ and store those with $as(\sum gh)>\varepsilon$ in a list L_{r1}.
4. Append to each $\sum gh\epsilon L_{r1}$ on either side Gestalten $f\epsilon L_0$ such that $\sum fgh$ and $\sum ghf$ and test for improvement of the assessment. If the longer row Gestalt is better replace it for the shorter one (following Desolneux's principle of *maximal meaningful Gestalt* [3]). Repeat until no better Gestalt is found.
5. Repeat from step 2 but with L_{m1} instead of L_0 and also with L_{r1} instead of L_0 forming L_{m2}, L_{r2}, etc. – until the scales of the found Gestalten become bigger than the image.

The benchmark images for the contest are mostly fairly small, so that there is not much sense in repeating the algorithm sketched above more often than two or three times. If a single mirror symmetry is the goal, the best Gestalten from the lists L_{ml} may be chosen as output. If multiple mirror symmetries are the goal, a threshold may be trained from the training set and the best Gestalten exceeding it may be chosen, and accordingly if row Gestalten (frieze symmetries) are the goal. An additional clustering step as described below in Section 2.3 fosters the recognition of global image symmetries.

The contest also contains lattice recognition. In the setting of Gestalt algebra as published in [9] this is possible by looking for rows of rows (i.e. $\sum\sum$-terms) where the outer orientation is roughly perpendicular to the inner ones. Unfortunately, this would lead to bad assessments if the rows have many members and the generator vectors are of comparable length in both directions. This could be alleviated by introducing a sixth dimension to the Gestalt manifold – an eccentricity. In an extended version of [9] planned for the PRIA journal this year this may be carried out. The same is planned for the inclusion of an operation for rotational symmetries. But

practical results cannot be given before the theory is settled. However, we include our results on the frieze data.

2.3. Accumulating clusters of Gestalten

Yet another Gestalt-law of [12] is *proximity*. In conjunction with *good continuation* this means that mirror-gestalts with roughly co-incident axes and in vicinity should be grouped into a cluster (prolonging the axis). In accordance with common practice in machine vision the axis is stored as homogenous 3-vector *a* choosing the image center as point of origin and scaling such that half of the minimum of image height and width is unity. Representatives are stored with the normal vector $(a_1, a_2)^T$ normed to 1. In this representation we calculate the distances.

Following the search according to Section 2.2 a greedy clustering is performed:

1. Pick the best assessed mirror Gestalt which is not already member of a cluster, initialize a new cluster with it, and add all mirror Gestalten closer to it than a predefined threshold.
2. Repeat 1 until a maximum number of clusters is achieved, or all mirror Gestalten are assigned to a cluster.

Each cluster is then evaluated by weighting with the assessments: $\sum as(g_i)$ where *i* runs over the members of the cluster. The best one (or best ones respectively) are transformed to the ground-truth format of the competition and thus form the output.

3. Experiments

The computational effort for the search as outlined in Section 2.2. is highly data dependent. It is not an exhaustive search, rather a heuristic greedy search. Thus it is fairly quick, and the benchmark with 32 images is processed in less than two minutes on state-of-the-art standard hardware.

The first step is always the extraction of primitives. Figure 1 depicts an example from the mirror symmetry contest with primitive SIFT-Gestalten overlaid as yellow circle (indicating position and scale) with a red radius indicating the most dominant orientation and blue radius indicating the second dominant orientation. Deliberately, we choose one of the larger images, as those better suit our ansatz.

Figure 2 shows some of the best assessed higher-order Gestalten from the lists of mirror objects of higher-orders (each in individual color). The largest Gestalt (in light green) probably meets approximately what a human would perceive. It is composed of the upper red and lower green mirror Gestalten which in turn are composed also as mirror Gestalten from SIFT-primitives. The line thickness

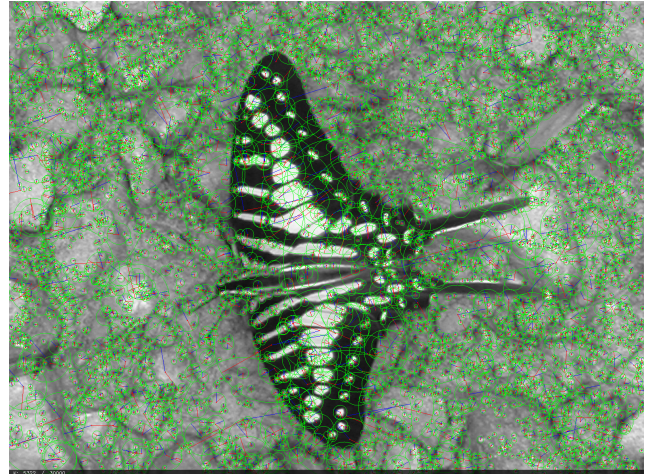of the drawing corresponds to the assessment of the Gestalt.

Figure 1: Example from the mirror test set with SIFT primitives overlaid
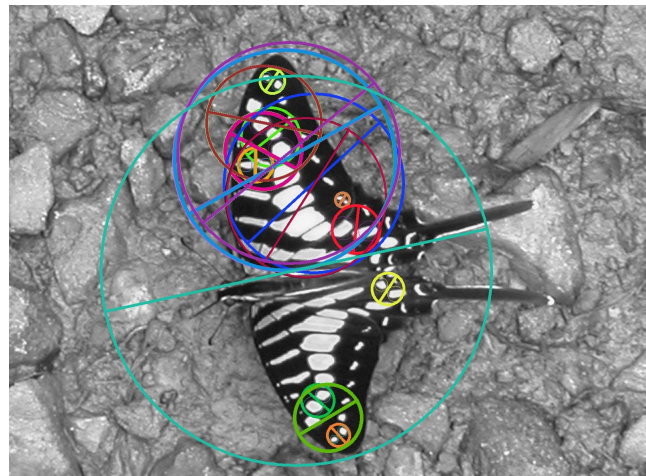

Figure 2: Some Gestalten on example image overlaid in individual colors

Figure 3 shows another rather successful example. However, this example also demonstrates that often the supporting region is underestimated as compared to human perception. And the clustering as indicated in Section 2.3 can only partially improve this. For the time being it is hard to explain why so many obviously symmetric parts are ignored. Often the primitives are missing due to a certain instability of the SIFT extractor (with its thresholds). Recall: The search process does not see the image, not even descriptors (e.g. histograms of orientations). It only gets position, scale, first dominant gradient orientation, and assessment. And this information is only given for above threshold primitives.

Quite often the mirror symmetry found by the Gestalt algebra approach is oriented just perpendicular to the one

preferred by human subjects (on approximately the same supporting region). Figure 4 shows such an example. It turns out that a posteriori human subjects would often agree with the found Gestalten as also being plausible. Sometimes we have the impression that in such cases humans tend to prefer vertical axes.
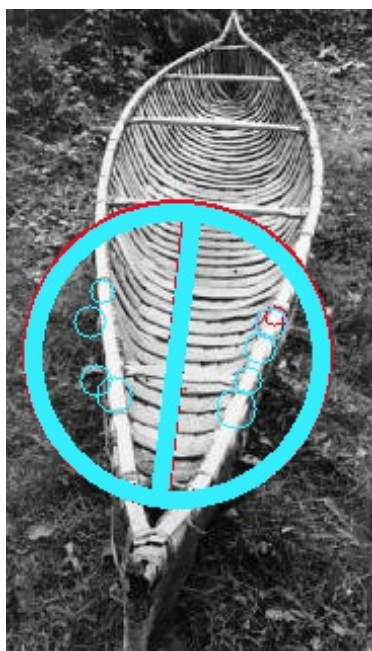

Figure 3: Best three |-Gestalten on an example image

Figure 4: Gestalten perpendicular to human preference

There are examples in the benchmark which display hierarchies of symmetries (such as displayed in Figure 5). Here with Gestalt ansatz the machine can recognize more than what was demanded for this competition. It can reproduce a hierarchy of symmetries of symmetries or friezes, or a frieze of symmetries, etc. of arbitrary depth.

This is what it was intended for, and on some examples it succeeds. All in all we have to admit that most of the competition images are too challenging for our method at its current status. For instance, in the example presented in Figure 6 the Gestalten look not much better than random. Still on such images humans immediately perceive meaningful symmetry! We stand in humbleness admiring the perceptive capabilities of humans.
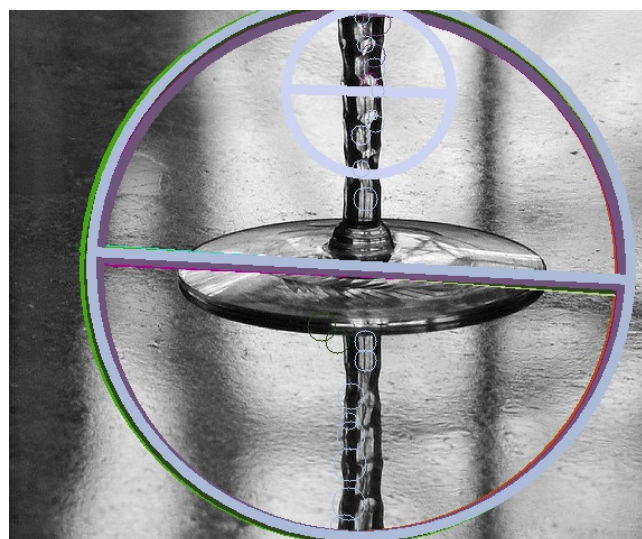

Figure 5: Higher order |-Gestalten

Figure 6: Complete failure on a more challenging example

For the mirror symmetry competition we hand in results before clustering and results after clustering and each for the single and multiple Symmetry sets respectively. For the frieze data there is no final clustering method implemented yet.

Figure 7 displays rows on an image from the frieze competition. Here we first displayed the ten best Gestalten using the same drawing procedure as for the |-Gestalten. So the diameter is drawn perpendicular to the orientation attribute of the Gestalt (like a mirror axis). From this of course the number of parts of such Gestalten cannot be seen. Therefore a drawing routine was implemented following the rationale of the groundtruth of the lattice symmetry contest. This separates the line on which the parts should ideally sit into n equally spaced segments and constructs squares on them. Each such square contains one

part of the ∑-Gestalt and also indicates its rough size. The yellow line does not circumscribe these squares, but draws the diagonals and separating lines in one zigzag.

Figure 7: Some ∑-Gestalten on an example image

In Figure 7 the best twenty ∑-Gestalten are overlaid over the result image from the Gestalt algebra output. It can be seen that most of them are very similar and in agreement with the correct vertical frequency. The number n is too large by one. The supporting region is again too small and a little displaced. All Gestalten sit on the right side, none on the left or in the middle of the tower. The first complete failure Gestalt comes with the best 20, but not with the best 10. Our experiments reveille, that the approach cannot handle strong projective or other geometric distortions.

4. Discussion

We participate in the competition, but we expect that other approaches will perform better in the distinct disciplines, as they may use more information – such as the colors of all pixels under concern – which is lost in our primitive extraction phase. See e.g. from the example in

Figures 1 and 2 that some of the correspondences set by the method do not perfectly satisfy and the vast majority of the information is not used at all. Still we are confident that future techniques may benefit from our definitions:

We imagine these techniques to use primitive extraction and Gestalt algebra terms together with an accumulator as a preliminary step. The best terms found in this way may be handed over to a verification step. This step may use any of the better performing other methods of the competition in respective sub-regions of the image. This should result in hierarchical symmetries of symmetries as they are used in the generic computer-graphics.

References

[1] M. Z. Aziz and B. Mertsching: An Attentional Approach for Perceptual Grouping of Spatially Distributed Patterns. In: F. Hamprecht et al. (eds.): Proc. 29th DAGM-Symposium, LNCS 4713, Springer, Berlin, 345-354, 2007.

[2] D. Cabrini Hauagge, N. Snavely. Image Matching using Local Symmetry Features. In Proc. Of Int. Conf. on Computer Vision and Pattern Recognition 2012.

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

[4] Finkenzeller, D., Bender, J. Semantic representation of complex building structures. Computer Graphics and Visualization (CGV 2008) - IADIS Multi Conference on Computer Science and Information Systems, Amsterdam, 2008.

[5] J.-H. Haunert: A symmetry detector for map generalization and urban-space analysis. ISPRS Journal of Photogrammetry and Remote Sensing, 74, 66-77, 2012.

[6] Y. Liu, H. Hel-Or, C. S. Kaplan, L. Van Gool, Computational Symmetry in Computer Vision and Computer Graphics. Foundations and Trends in Computer Graphics and Vision, 5 (1-2), 2010.

[7] R. Ma, J. Chen, and Z. Su. Mi-sift: mirror and inversion invariant generalization for sift descriptor. In Proc. ACM Image and Video Retrieval, 228–235, New York, NY, USA, ACM 83. 2010.

[8] E. Michaelsen, U. Stilla, U. Soergel, L. Doktorski. Extraction of building polygons from SAR images: Grouping and decision-level in the GESTALT system. Pattern Recognition Letters, 31(10), 1071-1076, 2010.

[9] E. Michaelsen, V. Yashina: Simple Gestalt Algebra. IMTA-4 2013, workshop along with VISAPP 2013 in Barcelona, 38-47, 2013.

[10] D. Reisfeld, H. Wolfson, Y. Yeshurun: Detection of Interest Points Using Symmetry, ICCV 1990, Osaka, 62-65, 1990.

[11] N. Ripperda: Grammar Based Façade Reconstruction Using rjMCMC. Photogrammetrie Fernerkundung Geoinformation, 2, 83-92, 2008.

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