DESIGN OF ORIENTATION ASSESSMENT FUNCTIONS FOR GESTALT-GROUPING UTILIZING LABELED SAMPLE-DATA

Eckart Michaelsen*, Jochen Meidow

Fraunhofer IOSB, Ettlingen, Germany - (eckart.michaelsen, jochen.meidow)@iosb.fraunhofer.de

KEY WORDS: perceptual grouping, Gestalt laws, parameter estimation

ABSTRACT:

Psychological evidence is given that perceptual grouping is an important help for various visual tasks. Object recognition and land use classification from remotely sensed imagery is an example. In machine vision, such a grouping process can be implemented by coding Gestalt laws such as proximity, symmetry, or good continuation. Since geometric relations are rarely fulfilled exactly, soft membership functions are utilized called Gestalt assessments. Hierarchical grouping is possible on increasing scales. Such an approach to hierarchical Gestalt grouping is modified in this paper. In its original form, the approach uses rather heuristic default assessment functions, which are a possible choice as long as no labeled example data are given. The assessment functions can be parameterized so as to improve the perceptual grouping, guiding it by the Gestalten salient to human perception. To this end, we use orientation statistics from the publicly available data set given for the ICCV symmetry recognition competition 2017. Also, with a particular recognition task at hand, labeled example data can serve as the desired foreground. Here we use the ground-truth layer for buildings of the Vaihingen benchmark of the ISPRS. A mixture distribution containing two von Mises-distributions and the uniform component for the clutter in the background is fitted using expectation maximization.

1. INTRODUCTION

Perceptual grouping along the Gestalt laws may have considerable improvement potential for various visual tasks, such as figure-ground organization or object recognition (Pizlo et al., 2014, Kanizsa, 1980). In (Michaelsen and Meidow, 2019) also remotely sensed examples are presented. A perceptual group is a finite set of *parts* that are seen together as one *aggregate*, i.e., the visual apparatus infers a common cause of the parts. Certain geometric relations must hold between the parts so that pre-attentive grouping occurs. These relations are known as the Gestalt laws. They include: *reflection symmetry*, i.e., the parts are mapped onto each other with respect to a mirror axis, *good continuation*, i.e., parts are repeated along a straight or at least smooth bending line, so that a frieze results in the discrete case, or a stripe in the continuous case, *proximity* — there is a tendency to group adjacent parts, etc. Such topics are not in the focus of machine vision or pattern recognition today. Yet, there is a large body of corresponding literature for which (Pizlo et al., 2014, Kanizsa, 1980, Desolneux et al., 2008) and (Michaelsen and Meidow, 2019) is only a very small sample. They contain references to a much wider scientific community, which agrees on the following points: Gestalt grouping is pre-attentive, it is fast, and it sets a third category of perception — beside perception based on training, and perception as an act of reasoning and logic inference.

The paper at hand can be seen on the interface between the first and the third category, i.e., between machine learning and perceptual grouping. The inherent parameters of perceptual grouping processes may well be subject to training or adjustment. In the last chapter of (Michaelsen and Meidow, 2019), we already augmented the Gestalt grouping approach by estimating optimal parameter settings for the assessment functions used in the grouping process. In particular, we addressed the orientation domain used in many grouping laws, such as *parallelism* or *orthogonality* or *reflection symmetry*. Several adjustment rationales are possible: *Mathematical inference* would derive parameters of orientation assessment functions from probabilistic assumptions about the perceived world, and the nature of the projection between the scene and the image, a *heuristic* adjustment would use common sense and trial and error on data at hand, and *parameter estimation* would assemble statistics on the mutual features of parts that are known to form a Gestalt-aggregate. In the paper at hand, we follow the latter path.

In this context, the grouping of gestalts is based on functions which assess the similarity and proximity of two or more entities considering various features, such as orientation or distance. Therefore, a distance function is required, which takes possibly correlated features of different kinds into consideration. Such a metric or distance function has to obey four axioms: non-negativity, identity of indiscernibles, symmetry, and subadditivity. In practice, metric learning algorithms ignore the condition of identity of indiscernibles and learn a pseudo-metric.

Thus the goal is to learn from examples the parameter values of a function that measures how similar or related two objects are. Corresponding distance functions such as the Mahalanobis distance should be unitless and scale-invariant. To learn the metrics, we exploit the statistics of labeled data sets. From such statistics, a parameterized density function for the relative orientation of parts of good gestalts can be estimated. This allows the design of better assessment functions for the similarity-in-orientation law. This idea has been proposed before for the determination of dominant orientations by fitting parametric distributions to the data (Pohl et al., 2017). However, here we assume the offset parameter of such distributions to be fixed, because for the Gestalt law *parallelism* the offset must be 0, and for *orthogonality* it must be π. Only the deviation parameters remain to be estimated.

Section 2 sets the context of hierarchical perceptual grouping. Section 3 constitutes the technical core of the paper, the continuous functions that define the model and the estimation of their

*Corresponding author

parameters. Section 4 gives the data and the resulting histograms and parameters. Finally, in Section 5, we discuss what has been achieved, conclude, and give an outline on the directions of related possible future research.

2. HIERARCHICAL GESTALT GROUPING – THE ROLE OF ASSESSMENT FUNCTIONS

Following (Michaelsen and Meidow, 2019), the domain for Gestalten g has five components: *location* $\boldsymbol{x}_g$ — a point in the ordinary 2D vector-space on the real field; *orientation* o_g — a real number representing an arc; *scale* s_g — a positive real number; *frequency* f_g with respect to rotational self-similarity — a non zero positive integer; and *assessment* a_g — a continuous measure for salience with values between 0 and 1. The latter can be interpreted as a fuzzy membership function. A Gestalt with assessment 1 is very meaningful and salient and a Gestalt with assessment 0 is meaningless.

Figure 1 shows the set of operations on the Gestalt domain, each standing for a specific visual grouping phenomenon. From top to down there are:

- *Reflection Symmetry*, aggregating a pair of Gestalten into one Gestalt.
- *Frieze Symmetry*, aggregating an n-tuple of Gestalten into one Gestalt.
- *Rotational Symmetry*, aggregating an n-tuple of Gestalten into one Gestalt.
- *Parallelism*, aggregating a pair of Gestalten into one Gestalt.
- *Good continuation and gap closing*, aggregating a finite set of Gestalten into one Gestalt.
- *Lattices*, aggregating an $n \times m$-tuple of Gestalten into one Gestalt.

On remotely sensed data, e.g., for building recognition or road extraction, parallelism and good continuation are most important. However, the other four aggregation laws can also contribute significantly.

For each aggregation, the laws of seeing are coded in the functions assessing the newly constructed aggregate. For example, *proximity* is coded as continuous function taking the location and scale features of two Gestalten as input and giving a result between 0 and 1. Proximity should not be confused with inverse distance. Objects positioned at the same location are not in proximity. The proximity assessment for such configuration should be zero, just as objects which are very far away from each other should have proximity assessment 0. The maximal proximity assessment 1 must be reached somewhere where the objects are adjacent to each other, i.e., where the distance equals the scale of the objects. (Michaelsen and Meidow, 2019) propose the use of

$$a_{\text{prox}}(g,h) = \exp\left(2 - \frac{d(\boldsymbol{x}_g, \boldsymbol{x}_h)}{\sqrt{s_g \cdot s_h}} - \frac{\sqrt{s_g \cdot s_h}}{d(\boldsymbol{x}_g, \boldsymbol{x}_h)}\right) \qquad (1)$$

for two gestalts g and h with the Euclidean distance $d(\boldsymbol{x}_g, \boldsymbol{x}_h)$. Another possible choice is a function that has the shape of the standard Rayleigh density but is normalized to maximum 1.

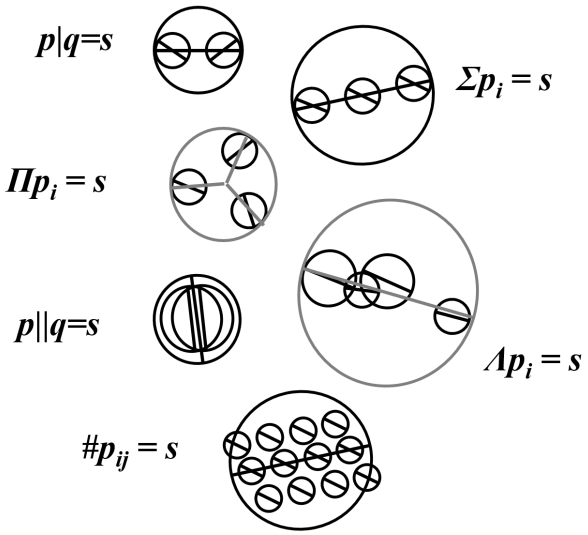

Figure 1. Operations on the Gestalt domain as given in (Michaelsen and Meidow, 2019), gestalts are displayed as circles with location at the center, scale as diameter, orientation as radius line, frequency as number of spokes, and assessment as gray tone

Proximity is combined with other laws, an important one being *similarity in orientation*. Recall orientations are given in a continuous additive group which is not a metric space. In (Michaelsen and Meidow, 2019) we proposed the use of functions that give 1 for no orientation difference, and 0 for maximal difference in orientation such as

$$a_{\text{ori}}(g,h) = \frac{1}{2} + \frac{1}{2} \cdot \cos(o_g - o_h)\,. \qquad (2)$$

Here g and h are again arbitrary gestalts, which must however have the same rotational frequency so that their orientations can be compared at all.

It is the intention of the paper at hand to improve this default choice using statistics on a labeled data set. From such statistics a parametrized density function for orientations of parts of good gestalts can be estimated. This allows the design of a better assessment function for the *similarity in orientation* law.

3. MODELING AND PARAMETER ESTIMATION

The histograms shown in Figures 2, 4, and 5 suggest that dominant orientations exist in our data. E.g., changes in building outlines can be modeled by a mixture of continuous parametric distributions on the unit circle. The peaks at 0 and π correspond to the omnipresent relations parallelism and orthogonality in man-made environments. Of course, further angles occur which are considered as background clutter since they hinder the grouping process.

3.1 Circular Distributions and Mixture Model

For the representation of orientation changes we utilzed the von Mises distribution which is in many respect the "natural" analogue on the circle of the normal distribution on the real line (Fisher, 1995). The probability density function reads

$$p(\alpha|\phi,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\alpha - \phi)\}\,, \qquad (3)$$

$0 \leq \alpha \leq 2\pi$, $0 \leq \kappa \leq \infty$, where $I_0(\kappa)$ is the modified Bessel function of order zero, ϕ is the mean direction, and κ is the so-called concentration parameter. As the concentration parameter κ approaches 0, the distribution converges to the uniform distribution; as κ approaches infinity, the distribution tends to the point distribution concentrated in the direction ϕ. Maximum likelihood estimates for the distribution parameters can be found in (Fisher, 1995) and (Best and Fisher, 1981).

Since we are dealing with orientation data, we have to consider distributions on the unit circle. For the statistical analysis, we transform the observed orientations $\alpha_i \bmod \pi$ by doubling them, estimate the distribution parameters, and back-transform the results. Furthermore, since we are dealing with orientation changes, we expect $\phi = 0$ for parallelism and prolongation and $\phi = \pi$ for orthogonality and therefore fix the parameter ϕ.

The background clutter, i.e., orientation changes not caused by parallelism or orthogonality, are modeled by the circular uniform distribution

$$p(\alpha) = \frac{1}{2\pi}, \qquad 0 \leq \alpha \leq 2\pi \tag{4}$$

i.e., all orientation changes are equiprobable.

Thus, we utilize a mixture of at least $D = 2$ von Mises distributions (3) and the uniform distribution (4), i.e.,

$$p(\alpha) = w_0 \cdot p(\alpha) + \sum_{d=1}^{D} w_d \cdot p(\alpha|\phi_d, \kappa_d), \quad \sum_{i=0}^{D} w_i = 1 \tag{5}$$

with $\phi_d \in \{0, \pi\}$ and unknown weights w_i for the components.

In the experiments, we check if further components are required to model the distributions appropriately.

3.2 Parameter Estimation and Model Selection

For the representation of the orientation changes we study two mixture models: The first one models orthogonality and parallelism with two von Mises distributions at 0 and π, plus a uniform distribution for the background noise ("2+1 model", see Figure 4). The second model sets two von Mises distributions in each case at 0 and π respectively to take additional clusters with less variation into account ("2×+1 model", see Figure 5). For the estimation of the distribution parameters and the weights of the mixture components, we apply the well-known expectation-maximization algorithm (Dempster et al., 1977) (EM). The iterative procedure can easily be initialized with equal weights for all components and moderate concentration parameters of $\kappa = 100$ for both von Mises distribution. For the model with five components ("2×+1 model"), we initialize with $\kappa_1 = 100$ for the narrow peak and $\kappa_2 = 10$ for the broader peak. For the latter case the algorithm converges after 2,690 iterations and provides the results depicted in Figure 5. For the time being, we visually inspect such results to assess the goodness-of-fit for the two models, bearing in mind their complexities.

4. RESULTS

We introduced mixture distributions with an in-between component between definite inliers and unrelated outliers in (Michaelsen and Meidow, 2014). The first experience with such estimations was made with a machine vision benchmark (Michaelsen and Meidow, 2019, Chapter 13). We recapitulate the accomplished evidence in Section 4.1. For the paper at hand, we augmented the investigation using a well-known remote sensing benchmark, and the resulting evidence is given in Section 4.2.

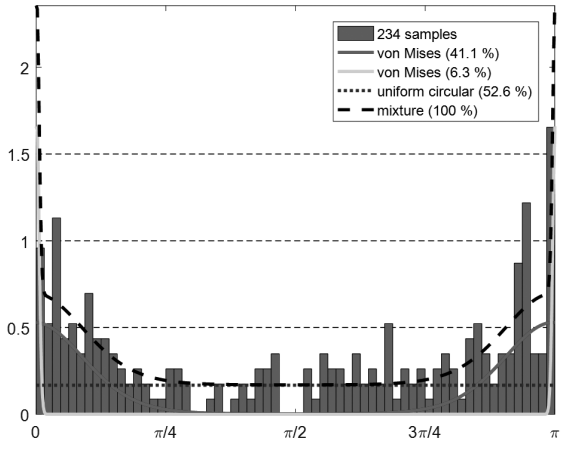

Figure 2. Histogram of observed orientations and estimated probability density functions of the mixture model.

4.1 Experiments – Utilizing the Frieze Competition of 2017

Along with the International Conference on Computer Vision 2017 in Venice, a research team from the Pennsylvania State University organized a competition on symmetry recognition (Funk et al., 2017). Among other categories, there also was frieze recognition. Fifty images were published with manually marked ground truth. In most of these images, one frieze is marked, in some images, more than one (but a small number), and in one none. For this work, we use at most one ground-truth per image, the first, so we have forty-nine ground truth frieze objects. In each of the forty-nine images, a set of primitive Gestalten is extracted using SLIC super-pixel segmentation (Achanta et al., 2012). Then an assessment-driven constant-false-alarm-rate search is performed on each set of such primitives. It searches for shallow-hierarchy gestalts using the laws for reflection symmetry and frieze formation of (Michaelsen and Meidow, 2019). The first criterion for the comparison of ground-truth frieze gestalts with automatically found gestalts is the number of parts that should exactly match. Then a gestalt distance is computed weighting location, scale, and orientation suitably. The best fitting Gestalt among the hierarchy 1 or 2 is selected, if it is closer than a suitable threshold.

If the best row Gestalt is found, the statistics of the orientations of the parts were centered to the mean orientation and recorded. On this statistic, we estimate the parameters of a mixture using the methods outlined above. More than half of the mass is uniformly distributed. There is a sharp narrow peak component that accounts for success examples where the orientations of the parts are very similar. Interestingly, between such outlier and inlier components, there exists an intermediate component, which is still narrower than the default assessment function. This result suggests that on these data such heuristic default function is suboptimal.

Instead, we estimated parameters for a corresponding mixture model on the statistics, using three components, one uniform for those parts where the parallelism law is more or less violated, one sharply peaked von Mises for those parts that obey the law, and a broader part. The latter captures in-between samples where the law of parallelism is weakened, e.g., due to perspective distortions, etc. The resulting mixture is displayed in Figure 2.

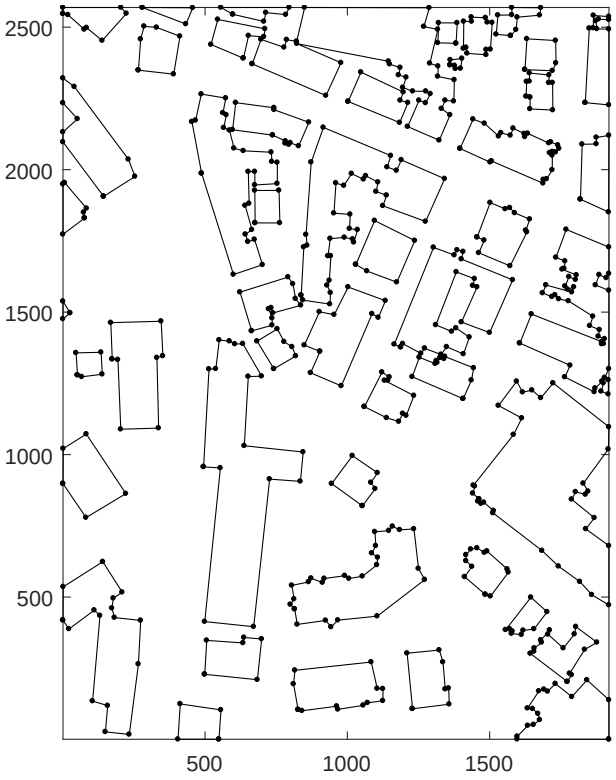

Figure 3. Extracted building outlines for the first index image.

4.2 Remotely Sensed Reference Data – the ISPRS-Vaihingen Benchmark

Different application domains may well yield different statistics. For this paper we made a comparative investigation using the Vaihingen data set provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (Cramer, 2010). Among other things, the data set provides 16 indexed images for common classes, e.g., buildings. The way corresponding to the work outlined above in Section 4.1 would be using the pseudo-color aerial images as input data. Then the standard perceptual grouping process would start. It would have to be modified so as to specialize in building recognition for instance. Then the building layer of the ground truth coming with the data would provide the target aggregates. The desired orientation statistics would result from the predecessors of the matching positives.

Instead, for simplicity, we used the building ground-truth images directly, assuming that their margin contours correspond well enough to the desired parts of the aerial image, i.e., those parts that should be preferably grouped in the image for the building recognition task. The corresponding Gestalt laws are of course *parallelism* and *orthogonality*.

For the vectorization of the buildings' outlines we initially trace the boundaries in the binaries images. Subsequently, the vertices of the resulting polygons are decimated. To do so, we consider the distance between a vertex and the straight line defined by the two adjacent vertices. We remove the vertices with a distance greater than five pixels in a greedy manner. Figure 3 shows the extracted outlines for the first ground truth image. In sum, we compiled 770 building outlines found in 16 images and obtained 6,635 polygon edges. For the specification of orientation changes, we determined the longest edge of each building can computed the angle between this edge and all other edges of this outline.

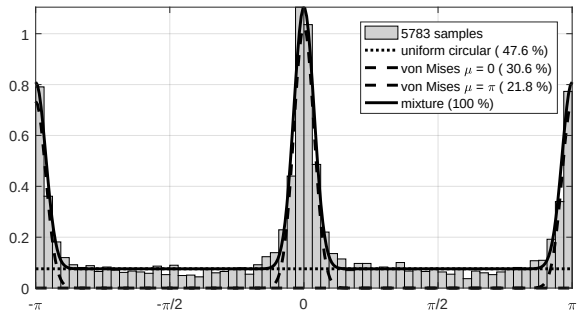

Figure 4. Histogram of the edge directions with 64 bins and estimated distributions of the mixture model with two von Mises distributions and the uniform circular distribution modeling the background clutter ("2+1 model").

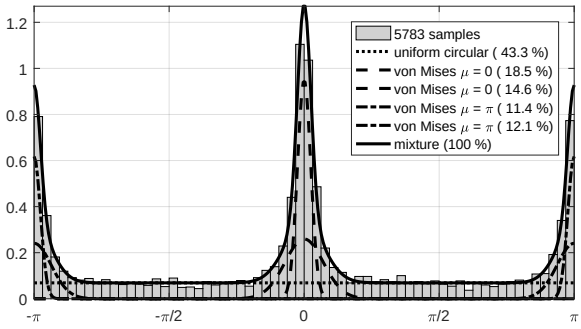

Figure 5. Histogram of the edge directions with 64 bins and estimated distributions of the mixture model with four von Mises distributions and the uniform circular distribution modeling the background clutter ("2×2+1 model").

Figure 4 shows the histogram of the directions with 64 bins and the estimated distribution of the mixture model with two von Mises distributions capturing parallelism and orthogonality and the circular uniform distribution ("2+1 model"). The background noise comprises 47.6 % of all observations.

The goodness-of-fit appears to be sub-optimal at the flanks of the peaks. Therefore, we introduced two additional von Mises distribution to model further orientation changes with a larger variation. Figure 5 shows the corresponding result.

Given an estimated concentration parameter, the circular standard deviation $\sigma_c = \sqrt{-2\log\rho}$ with $\rho = I_1(\kappa)/I_0(\kappa)$ can be computed to specify the variation of the orientation changes. For the mixture model with five components we estimated $\kappa_1 \approx 165$ for the narrow peaks and $\kappa_2 \approx 20$ for the more broader peaks. Thus, we obtain $\sigma_1 = 4.5^\circ$ and $\sigma_2 = 13.0^\circ$ which can be used to specify metrics used in distance functions.

5. DISCUSSION AND CONCLUSION

The very idea of hierarchical Gestalt grouping rests on its universal claim to be valid independent of any learning data — representative always only for a portion of the world. Such perceptual grouping should be already working to some degree even with inputs of a kind never seen before. However, some parameters in the system may have initial default values turning out to be sub-optimal. Gestaltists never denied the merit of machine learning (Sarkar and Boyer, 1994). Always, better results can be achieved if some of these parameter values are trained using suitable data. In this paper, the focus was on the orientation similarity

assessment. It turns out that additive mixture models are required to capture what is encountered in the orientations of the parts of true positive Gestalten. It is essential to utilize data not selected and labeled by the authors of the system themselves.

For this paper we acknowledge the work provided by the team of the Pennsylvania State University as well as the ISPRS benchmarking services. Comparing Figures 2 and 5 we conclude that the outcome is qualitatively similar as well for "images in the wild" as for remotely sensed imagery: Two inlier components, one very narrow and one more liberal, were modeled by von Mises-distributions. The third component is a uniformly distributed component. This copes for the cases when aggregates do well suit the ground-truth, that are made from parts whose orientations are not similar at all. There are, however, significant differences in the obtained parameters: The inlier component resulting from the symmetry benchmark is very peaked and has only a comparatively small weight, while the in-between component is quite broad and has a large weight. In contrast, on the remotely sensed data these two components turn out closer to each other as well in width as in weight.

There remain serious issues to be investigated and discussed.

- We are well aware that for mixture models the maximum-likelihood estimation is not consistent, i.e., in principle, for any setting, there is no guarantee to find the right mixture when the number of observations approaches infinity. Our visual inspection on the fit of the curves with the histograms can only serve as a preliminary result.

- As long as assessment functions are only used to compare or sort with respect to the fitness of configurations of the same type, e.g., only compare how well parallelism is given, nothing has been gained. Any similar monotone function will do the same, including the previous default assessments. Only if the assessment functions are combined, e.g., in the styles listed in (Michaelsen and Meidow, 2019), a benefit will appear along lines given here near the end of Section 1. E.g., already if configurations are assessed with respect to parallelism *and* proximity, a mutually sound assessment for both will be required. So the paper at hand is a step in this rationale. However, such combinations will also require the investigation of correlations between parallelism and proximity, and the further goal is of course the propagation of such assessments through larger hierarchies.

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel, methods. *Transactions on Pattern Analysis and Machine Intelligence*, 34(11), 2274–2281.

Best, D.J., Fisher, N.I., 1981. The BIAS of the maximum likelihood estimators of the von Mises-Fisher concentration parameters: The BIAS of the maximum likelihood estimators. *Communications in Statistics-Simulation and Computation*, 10 (5), 493–502.

Cramer, M., 2010. The DGPF test on digital aerial camera evaluation – Overview and test design. *Photogrammetrie – Fernerkundung – Geoinformation*, 2, 73–82.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39, 1–38.

Desolneux, A., Moisan, L., Morel, J.-M., 2008. *From Gestalt Theory to Image Analysis: A Probabilistic Approach*. Springer.

Fisher, N.I., 1995. *Statistical analysis of circular data*. Cambridge University Press.

Funk, C., Lee, S., Oswald, M. R., Tsokas, S., Shen, W., Cohen, A., Dickinson, S., Liu, Y., 2017. 2017 ICCV challenge: Detecting symmetry in the wild. In: *International Conference on Computer Vision 2017, Workshops*.

Kanizsa, G., 1980. *Grammatica del vedere. Saggi su percezione e gestalt*. Il Mulino.

Michaelsen, E., Meidow, J., 2014. Stochastic reasoning for structural pattern recognition: An example from image-based UAV navigation. *Pattern Recognition*, 47(8), 2732–2744.

Michaelsen, E., Meidow, J., 2019. *Hierarchical Perceptual Grouping for Object Recognition – Theoretical Views and Gestalt Law Applications*. Advances in Computer Vision andPattern Recognition, 1 edn, Springer International Publishing.

Pizlo, Z., Li, Y., Sawada, T., Steinman, R.M., 2014. *Making a Machine that Sees Like Us*. Oxford University Press.

Pohl, M., Meidow, J., Bulatov, D., 2017. Simplification of polygonal chains by enforcing few distinctive edge directions. In: P. Sharma and F. Bianchi (eds), *Scandinavian Conference on Image Analysis (SCIA)*, Lecture Notes in Computer Sciences, 10270, 1–12.

Sarkar, S., Boyer, K.L.,1994. *Computing Perceptual Organization in Computer Vision*. World Scientific.