The Gestalt Principle Collinearity and the Multi–Modal Statistics of Natural Image Sequences

Norbert Krüger and Florentin Wörgötter
University of Stirling, Scotland, norbert{worgott}@cn.stir.ac.uk

Abstract

We investigate the multi-modal statistics of natural image sequences looking at the modalities orientation, color, optic flow and contrast transition. We can show that the Gestalt law collinearity is measurable in visual data as a second order event of local line segment detectors. Furthermore, we can show that the statistical interdependencies corresponding to the Gestalt law collinearity increase significantly when we look not at orientation only but also at other modalities in addition to orientation. By these investigation, the attempt to model the application of Gestalt laws in computer vision systems based on statistical measurements (as suggested recently by some researchers ([10, 7, 15, 23])) gets further support. The results in this paper also suggest to formulate Gestalt principles in artificial vision systems in a multi-modal way. We discuss the potential usage of statistical interdependencies measured in this work within artificial visual systems and show first results.

Keywords: Grouping, Modality Integration, Statistics of Natural Images

1 Introduction

A large amount of research has been focused on the usage of Gestalt laws in computer vision systems (overviews are given in [22, 21]). The most often applied and also the most dominant Gestalt principle in natural images is collinearity [7, 15]. Collinearity can be exploited to achieve more robust feature extraction in different domains, such as, edge detection (see, e.g., [11, 12]) or stereo estimation [5, 21]. In most applications in artificial visual systems, the relation between features, i.e., the applied Gestalt principle, has been defined heuristically. Mostly, explicit models of feature interaction have been applied, connected with the introduction of parameters to be estimated beforehand, a problem recognized as extremely awkward in computer vision. Recently, Geisler et al [10] introduced the idea to overcome heuristic and explicit models by relating feature interaction to the statistics of natural images. The feasibility of this approach becomes strong support from the *measurable interdependencies* of features in visual scenes as performed here and in some recent work [15, 7, 10].

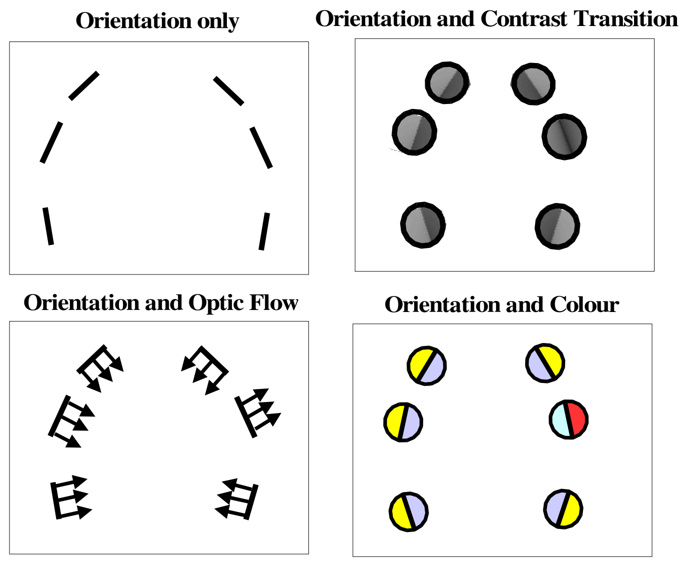

Figure 1. Grouping of entities becomes intensified (left triple) or weakened (right triple) by using additional modalities: Since the visual entities are not only collinear but also show similarity in another modality their grouping becomes more likely.

Many publications have addressed the question of efficient coding of visual information and its relation to the statistics of natural images (excellent overviews are given in [26, 24]). While many publications were concerned with the statistics on the pixel level and the derivation of filters from natural images by coding principles (see, e.g. [19, 2]), recently statistical investigation in the position–orientation feature space (i.e., ssstatistics of local line detector responses) have been performed (see, e.g., [15, 7, 10]) and have addressed the representation of Gestalt principles in this feature space. This relates to an old idea. Decades ago,

Figure 2. Top: Images of the data set. Bottom: 3 Images from a sequence.

Brunswick and Kamiya [4] first had stated that Gestalt principles should be related to the statistics of the natural world. Unfortunately the limited computational power at this time made it difficult to quantitatively support this statement. However, recently, the strong prevalence of collinearity in natural images could be measured first by [15] and [7]. These results have been confirmed and extended by [23, 10].

Interestingly, the occurrence of illusionary contour processing (in which the Gestalt law 'collinearity' is also tightly involved) develops at a late stage (after approximately 6 months) during the development of the human visual system (see [3] and [17]). This late occurrence of the above mentioned mechanisms suggets that those depend on visual experience about the underlying structures in visual data. This also suggests a formalization of Gestalt laws in artificial systems depending on statistical measurements and thereby replacing heursitics by statistics.

In the human visual system beside local orientation also other modalities such as color and optic flow are computed (see, e.g., [9]). All these low level processes face the problem of an extremely high degree of vagueness and uncertainty [1]. However, the human visual systems acquires visual representations which allow for actions with high precision and certainty within the 3D world under rather uncontrolled conditions. The human visual system can achieve the necessary certainty and completeness by *integrating* visual information across modalities (see, e.g., ([20, 13]).

Also Gestalt principles are affected by multiple modalities. For example, figure 1 shows how collinearity can be intensified by the different modalities contrast transition, optic flow and color. This paper addresses statistics of natural images in these modalities. As a main result we find that statistical interdependencies corresponding to the Gestalt law "collinearity" in visual scenes become significantly stronger when multiple modalities are taken into account (see section 3). Furthermore, we show first results how instead of using explicit rules for collinearity a simple criterion based on our measurements can be used to code collinearity within an artificial systems (see section 4).

2 Feature Processing

In the work presented here we address the multi–modal statistics of natural images. We start from a feature space (see also figure 3) containing the following sub-modalities:

Orientation: We compute local orientation o (and local phase p) by the filter [8].

Contrast Transition: The contrast transition of the signal is coded in the phase p of the same filter. The phase at a local maximum can be used to interpret the kind of contrast transition at this maximum [14], e.g., a phase of $\frac{\pi}{2}$ corresponds to a dark–bright edge, while a phase of 0 corresponds to a bright line on dark background. The continuum of contrast transition at an intrinsic one-dimensional signal patch can be expressed by the continuum of phases.

Color: Color is processed by integrating over image patches in coincidence with their edge structure (i.e., integrating over the left and right side of the edge separately). Hence, we represent color by the two tuples $\vec{c}_l(c^l_r, c^l_g, c^l_b), \vec{c}_r = (c^r_r, c^r_g, c^r_b)$ representing the color in RGB space on the left and right side of the edge. Note that our coding of color information does not contain any phase information since the tupels (c_r, c_g, c_b) are normalized to norm one, i.e., $c_r + c_g + c_b = 1$ with $c_i \geq 0$.

Optic Flow: Local displacements $\vec{o} = (f_1, f_2)$ are computed by the well known optical flow technique [18].

All modalities are extracted from a local image patch[1], resembling to columns in V1 responsible for a certain retina patch. The output is a local interpretation of the image patch by semantic properties (such as orientation and displacement).

For our statistics we use 9 Image sequences with a total of 42 images of size 512x512 (18 images) and 384x288 (24 images). The image sequences contain variations caused by object motion as well as camera motion (see figure 2). There is a total of 3900 feature vectors in the Data set (approximately 2600 from the outdoor images) and the statistic is based on 1555548 second order comparisons.

[1] In our statistical measurements we only use image patches corresponding to intrinsically one-dimensional signals (see [26]) since orientation is reasonably defined for these image patches only.

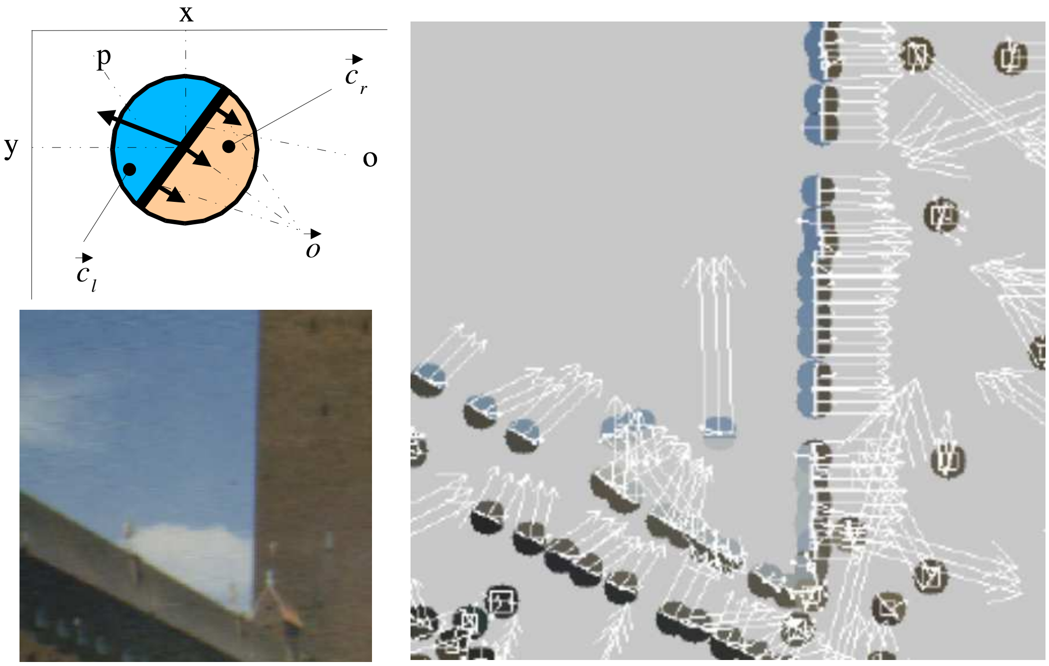

Figure 3. Top left: **Schematic representation of a basic feature vector.** Bottom left: **Frame in an image (from image sequence shown in figure 2 (bottom)).** Right: **Extracted Feature vectors.**

3 Statistical Interdependencies in Image Sequences

We measure statistical interdependencies by the so called 'Gestalt coefficient' which is defined by the ratio of the likelihood of an event e^1 given another event e^2 and the likelihood of the event e^1:

$$G(e^1, e^2) = \frac{P(e^1|e^2)}{P(e^1).} \tag{1}$$

For the modeling of feature interaction a high Gestalt coefficient is helpful since it indicates the modification of likelihood of the event e^1 depending on other events. A Gestalt coefficient of one says, that the event e^2 does not influence the likelihood of the occurrence of the event e^1. A value smaller than one indicates a negative dependency: the occurrence of the event e^2 reduces the likelihood that e^1 occurs. A value larger than one indicates a positive dependency: the occurrence of the event e^2 increases the likelihood that e^1 occurs. The Gestalt coefficient is illustrated in figure 4. Further details can be found in [16].

3.1 Second Order Relations Statistics of Natural Images

Here we not only perform statistics in the position–orientation domain (as done by [15, 7, 10]) but we go one step further by investigating the second order relations of events in the *multi–modal* feature space

$$e = ((x, y), o, p, ((c_r^l, c_g^l, c_b^l), (c_r^r, c_g^r, c_b^r)), (f_1, f_2))$$

described above.

The distribution of the position (x, y) of entities in the image space is approximately isotropic. However, the distribution of orientations in the extracted entities is non–isotropic (see, e.g., [15]) with a significantly higher density for vertical and horizontal orientation than for diagonal orientation. Here, for the sake of simplicity, we want to neglect this anisotropy (for a detailed discussion see [15]). Therefore, in our investigation of second order relations we apply a transformation to the coordinate system such that the entity e^2 in the tuple (e^1, e^2) has zero orientation and zero position.

In our measurements we collect second order events in bins defined by small patches in the (x_1, x_2)–space and by regions in the modality–spaces defined by the metrics defined for each modality (for details see [16]). Figure 5 shows the Gestalt coefficient for equidistantly separated bins (one bin corresponds to a square of 10×10 pixels and an angle of $\frac{\pi}{8}$ rad). As already been shown in [15, 10] collinearity can be detected as significant second order relation as a ridge along the x-axis in the surface plot for $\Delta o = 0$ in figure 5e. Also parallelism is detectable as an offset of this surface (for

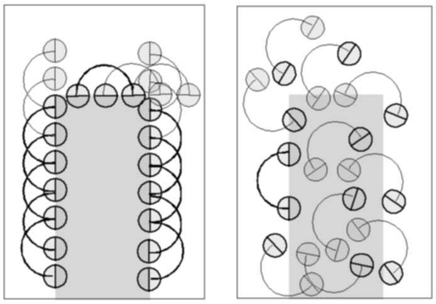

Figure 4. Let the second order event e^1 be: "occurrence of collinear line segments two units away from an existing line segment e^2". Left: Computation of $P(e^1|e^2)$. All possible occurrences of events e^1 in the image are shown. Bold arcs represent real occurrences of the specific second order relations e^1 whereas arcs in general represent possible occurrences of e^1. In this image we have 17 possible occurrences of collinear line segments two units away from an existing line segment e^2 and 11 real occurrences. Therefore we have $P(e^1|e^2) = 11/17 = 0.64$. Right: Approximation of the probability $P(e^1)$ by a Monte Carlo method. Entities e^2 (bold) are placed randomly in the image and the presence of the event 'occurrence of collinear line segments two units apart of e^2' is evaluated. (In our measurements we used more than a 500000 samples for the estimation of $P(e^1)$). Only in 1 of 11 possible cases this event takes place (bold arc). Therefore we have $P(e^1) = 1/11 = 0.09$ and the Gestalt coefficient for the second order relation is $G(e^1|e^2) = 0.64/0.09 = 7.1$.

details see [16] and [15]). A Gestalt coefficient significantly above one can also be detected for small orientation differences (figure 5d,f, i.e., $\Delta o = -\frac{\pi}{8}$ and $\Delta o = \frac{\pi}{8}$) corresponding to curved lines. The general shape of surfaces is similar in all following measurements concerned with additional modalities: we find a ridge corresponding to collinearity and an offset corresponding to parallelism and a Gestalt coefficient close to one for all larger orientation differences. Therefore, in the following we will only look at the surface plots for equal orientation $\Delta o = 0$.

These result shows that Gestalt laws are reflected in the statistics of natural images: Collinearity and parallelism are significant second order events of visual low level filters (see also [15]).

3.2 Pronounced Interdependencies by using additional Modalities

Now we can look at the Gestalt coefficient when we also take into account the modalities contrast transition, optic flow and color.

Orientation and Contrast Transition: We say two events $((x_1, x_2), o)$ and $((x'_1, x'_2), o')$ have contrast transition (i.e., 'similar phase') when $d(p, p') < t^{p+}$. t^{p+} is defined such that only 10% of the comparisons $d(p, p')$ in the data set are smaller than t^p. The metric for phase and all other modalities is defined in precisely in [16].

Figure 6b shows the Gestalt coefficient for the events 'similar orientation and similar contrast transition'. In figure 7 the Gestalt coefficient along the x-axes in the surface plot of figure 6 (i.e., the collinearity ridge) is shown. The Gestalt coefficient on the x-axes corresponds to the 'collinearity' ridge.

The first column of each group of 8 columns in figure 7 represents the Gestalt coefficient when we look at similar orientation only, while the second columns represent the Gestalt coefficient when we look at similar orientation and similar phase. We see a significant increase of the Gestalt coefficient compared to the case when we look at orientation only for collinearity.

Orientation and Optic Flow: The corresponding surface plot is shown in figure 6c and the slice corresponding to the collinearity ridge is shown as the third columns in figure 7. An even more pronounced increase of inferential power for collinearity can be detected.

Orientation and Color: Analogously, we define that two events have 'similar color structure'. The corresponding surface plot is shown in figure 6d and the slice corresponding to the collinearity ridge is shown the fourth columns in figure 7.

Multiple additional Modalities: Figure 6 shows the surface for similar orientation, phase and optic flow (figure 6e); similar orientation, phase and color (figure 6f) and similar orientation, optic flow and color (figure 6g). The slices corresponding to collinearity are shown in the fifth to seventh columns in figure 7. We can see that the the Gestalt coefficient for collinear line segments increases significantly. Most distinctly for the combination optic flow and color (seventh column). Finally we can look at the Gestalt coefficient when we take all three modalities into account. Figure 6h and the eighth column in figure 7 shows the results. Again an increase of the Gestalt coefficient compared to the case when we look at only two additional modalities can be achieved.

This result shows that assuming a line segment with a certain contrast transition does exist in an image it not only that the likelihood for the existence of a collinear line segment increases but that it also more likely that it has similar contrast transition.

4 Conclusion and Discussion

In this paper we have addressed the statistics of local oriented line segments derived from natural scenes. We could validate that the Gestalt law collinearity is

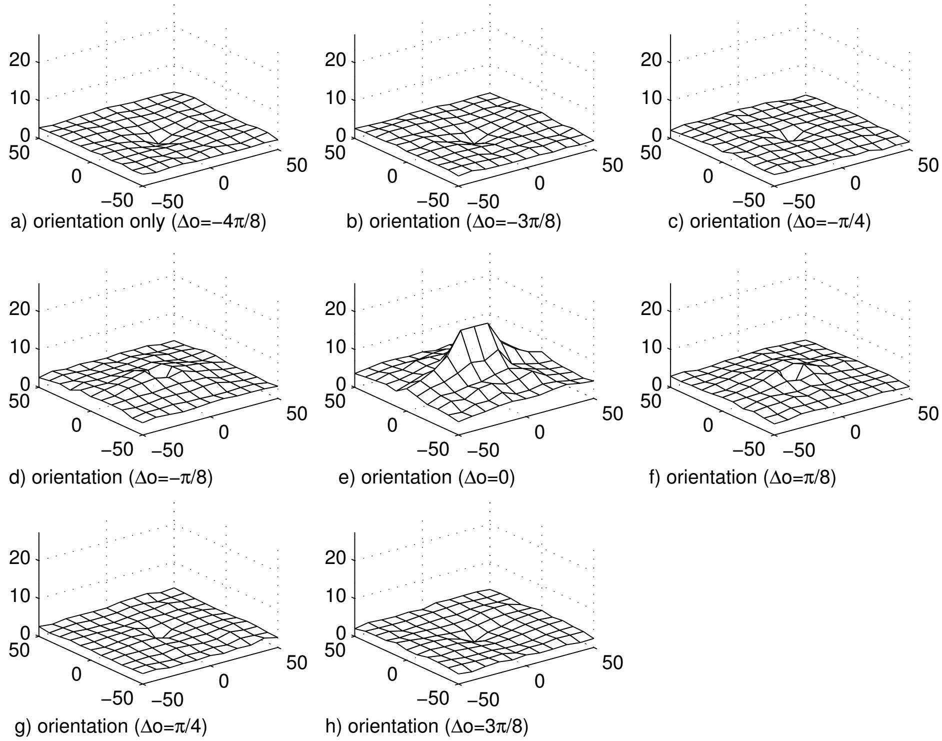

Figure 5. The Gestalt coefficient for differences in position from -50 to 50 pixel in x– and y– direction when orientation only is regarded. In a) the differecne of orientation of the line segments is $\frac{\pi}{2}$ (the line segments are orhtogonal) while in e) the difference of orientation is 0, i.e., the line segments have same orientation. The b), c), d) represent orientation difference between $\frac{\pi}{2}$ and 0. Note that the Gestalt coefficient for position (0,0) and $\Delta o = 0$ is set to the maximum of the surface for better display. The Gestalt coefficient is not interesting at this position, since e^1 and e^2 are identical

measurable in natural images as a second order event in the statistics of a local line segment detectors. Furthermore, by adding information on contrast transition, color, and optic flow we could show that statistical interdependencies in the orientation–position domain corresponding to collinearity become significantly stronger if one additional modality is taken into account. The statistical interdependencies increase again when multiple modalities are added. Essentially it seems that visual information bears a high degree of intrinsic redundancy that potentially can be used to reduce the ambiguity of local feature processing.

The measured multi–modal interdependencies can be used for the formalization of Gestalt principles in artificial systems in a probabilistic framework (see, e.g., [10, 6]). The claim is that intrinsic regularities in visual data allow for predictions which can be used to stabilize locally extracted information. To justify this, we want to address the issue of predictions in more detail now. At least three kinds of predictions which all make use of the Gestalt coefficient measured here, can be developed immediately:

Existence: The existence of an entity e^1 becomes more likely when an entity e^2 does exist and $G(e^1, e^2)$ is high. This kind of prediction can be used, e.g., to stabilize the system's confidence for the existence of e^1 in case that there is only moderate evidence (for the existence of e^1) from local operations, i.e. at the basic feature extraction stage (see figure 8a).

Grouping: Grouping can be achieved by assembling those extracted feature vectors into a group for which high statistical interdependencies have been measured, i.e., in case that there is good evidence from local operations that e^1 and e^2 do exist and $G(e^1, e^2)$ is high they can be assumed to belong to the same group (see figure 8b, first arrow). As already suggested in [10], these feature assemblies can then be enlarged and stabilized by using the transitivity relation, i.e., if e^1 and e^2 belong to a common feature group and e^2 and e^3 belong to a common feature group than e^1 and e^3 belong to a common group as well (see figure 8b, second arrow). Note that in this approach grouping does not require explicitly defined higher feature constellation but is performed *dynamically*, i.e. features become as-

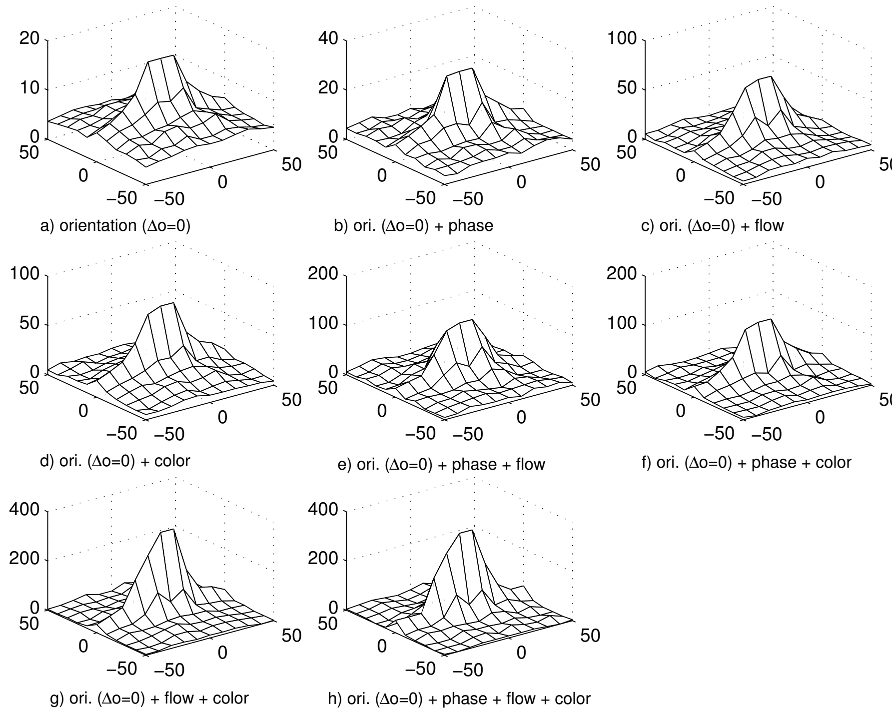

Figure 6. The Gestalt coefficient for $\Delta o = 0$ and all possible combination of modalities.

sembled according to the current input [25].

Feature Disambiguation: In case that entities $e^1, e^2, \ldots$ belong to the same group it can be assumed that the semantic properties of the entities are influenced by all entities belonging to the same group (see figure 8c and 8d). For different modalities such interactions have to formalized. Since metrics can be defined for all modalities we can speak about similarity in a modality. One possible formalization of an interaction rule can be: If the entities $e^1, e^2, \ldots$ belong to the same group and are similar in a certain modality (e.g., have similar color structure) we can define the value in this modality (e.g., the RGB-vector $\vec{c}_l = (c_r^l, c_g^l, c_b^l)$) for each entity as the (weighted) average over the values of all entities in the group in this modality. In addition, it may be that also specifics of the modalities have to be taken into account (e.g., the aperture problem for optic flow).

The formalization of interaction schemes based on the measured Gestalt coefficient are addressed in our current research. Here, we want to describe two examples: A process of self–emergence of feature constellations and low–contrast edge detection. In both cases only a simple criterion based on the Gestalt coefficient is applied to realize the collinear relation.

Self–Emergence of Feature Constellations: The need of entities going beyond local oriented edges is widely accepted in computer vision systems across a wide range of different viewpoints. Their role is to extract from the complex distribution of pixels in an image patch (or an image patch sequence) a sparse and higher semantical representation which enables rich predictions across modalities, spatial distances and frames. Accordingly, they consist of groups of early visual features (such as local edges).

These higher feature constellations have been already applied in artificial systems but were needed to be defined heuristically. By using a link criterion based on the Gestalt coefficient (stating that there exist a link when the Gestalt coefficient is high) and the transitivity relation (if two pairs of entities are linked then all entities have to be linked) we are able to define a process in which groups of local entities do self emerge. Since the link relation defines an equivalent relation on the space of entities this space can be separated in disjunct linked groups which can be found by a simple algorithm. In figure 9 (left) such groups are labeled by same color at the center of the feature patches.

Detection of low Contrast Edges: Once the groups have self–emerged they can be used to detect low contrast edges and reduce falsely detected edges caused by structural noise by combining local confidence and contextual confidence selected within the group an entity belongs to. In figure 9b all features above a certain threshold are displayed with a filled circle. Note the detection of the low contrast edge (figure 9b, horizontal ellipse) when applying grouping based on the Gestalt coefficient and the reduction of non meaningful features (vertical ellipse) without grouping. The extraction of features using only local information is shown in figure

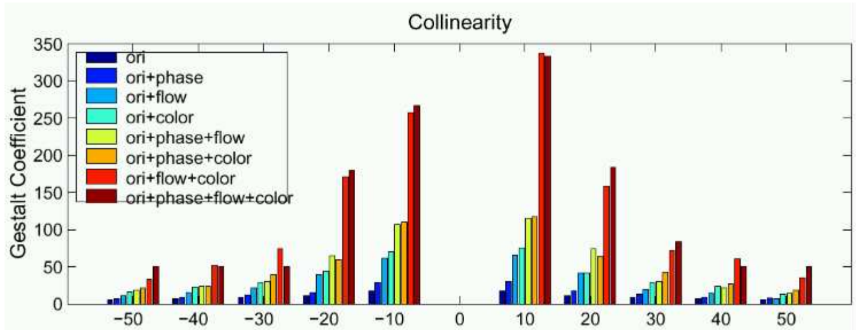

Figure 7. The Gestalt coefficient for collinear feature vectors for all combinations of modalities. For (0,0) the Gestalt coefficient is not shown, since e^1 and e^2 would be identical.

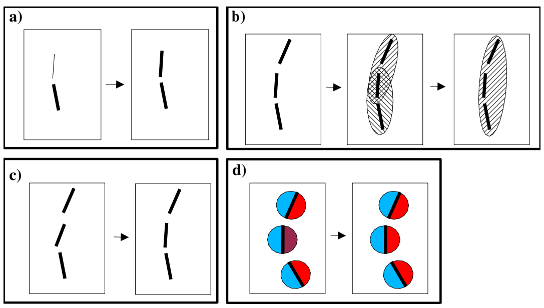

Figure 8. Schematic description of different kind of predictions. a) Existence: **The low likelihood for the existence of the entity e^1 (indicated as thin line) increases because of the better consolidated existence of e^2 (indicated as bold line) and the high Gestalt coefficient $G(e^1, e^2)$ for such a constellation.** b) Grouping: **The concurrent existence of collinear entities (i.e., entities with high $G(e^1, e^2)$) leads to grouping into tuples and then (by the transitivity rule) to larger feature assemblies.** c+d) Feature Disambiguation: **Orientation correction (c) and color correction (d) across entities assembled within one group.**

9c.

The exact formalization of grouping and feature disambigution based on sthe statistical measurements explained here is part of our current research and will be described in a forecoming paper.

References

[1] J. Aloimonos and D. Shulman. *Integration of Visual Modules — An extension of the Marr Paradigm.* Academic Press, London, 1989.

[2] A.J. Bell and T. Sejnowski. Edges are the 'independent components'of natural scenes. *Advances in Neural Information Processing Systems*, 9, 1996.

[3] B.I. Bertenthal, J.J. Campos, and M.M. Haith. Development of visual organisation: The perception of subjective contours. *Child Development*, 51(4):1072–1080, 1980.

[4] E. Brunswick and J. Kamiya. Ecological cue–validity of 'proximity' and of other gestalt factors. *American Journal of Psychologie*, LXVI:20–32, 1953.

[5] R.C.K. Chung and R. Nevatia. Use of monucular groupings and occlusion analysis in a hierarchical stereo system. *CVPR*, 1991.

[6] A. Desolneux, L. Moisan, and J.M. Morel. Edge detection by the Helmholtz principle. *JMIV*, 14(3):271–284, 2001.

[7] H. Elder and R.M. Goldberg. Inferential reliability of contour grouping cues in natural images. *Perception Supplement*, 27, 1998.

[8] M. Felsberg and G. Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 41(12), 2001.

[9] M.S. Gazzaniga. *The cognitive Neuroscience.* MIT Press, 1995.

[10] W.S. Geisler, J.S. Perry, B.J. Super, and D.P. Gallogly. Edge co–occurrence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724, 2001.

[11] G. Guy and G. Medioni. Inferring global perceptual contours from local features. *International Journal of Computer Vision*, 20:113–133, 1996.

[12] F. Heitger, R. von der Heydt, E. Peterhans, L. Rosenthaler, and O. Kübler. Simulation of neural contour meachnisms: representing anomalous contours. *Image and Vision Computing*, 16:407–421, 1998.

[13] D.D. Hoffman, editor. *Visual Intelligence: How we create what we see.* W.W. Norton and Company, 1980.

[14] P. Kovesi. Image features from phase congruency. *Videre: Journal of Computer Vision Research*, 1(3), 1999.

[15] N. Krüger. Collinearity and parallelism are statistically significant second order relations of complex cell responses. *Neural Processing Letters*, 8(2), 1998.

[16] N. Krüger and F. Wörgötter. Multi modal estimation of collinearity and parallelism in natural image sequences. *Submitted.*

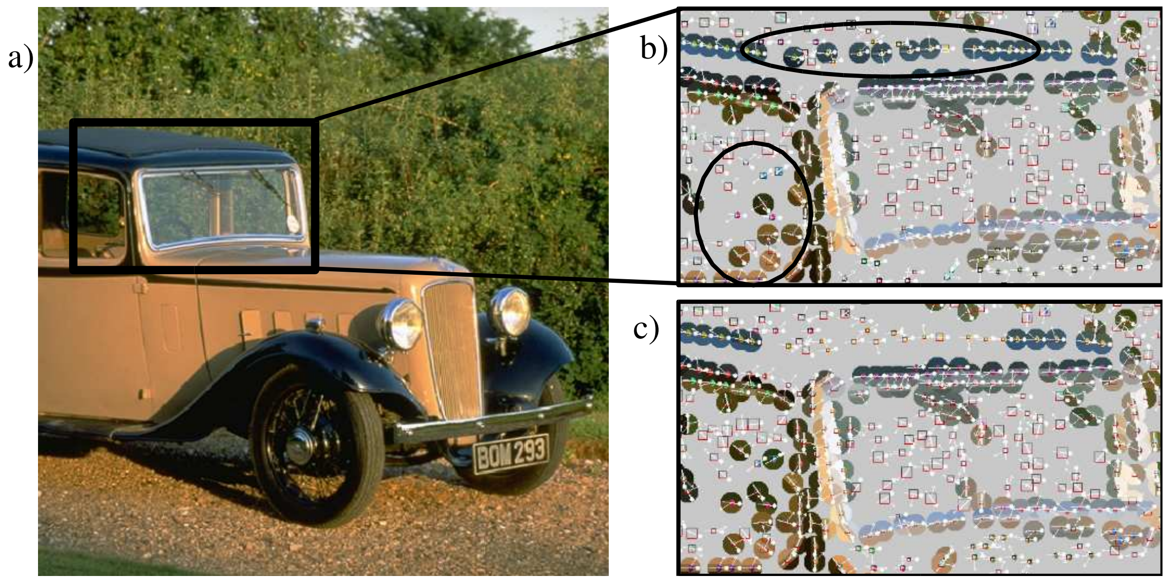

Figure 9. a: Image of a car. b: Extraction of features with grouping based on the Gestalt coefficient. c: Feature extraction without grouping. Note that this is a color figure. Its meaning is only fully understandable using color information.

[17] N. Krüger and F. Wörgötter. Different degree of genetical prestructuring in the ontogenesis of visual abilities based on deterministic and statistical regularities. *Proceedings of the 'Workshop On Growing up Artifacts that Live' SAB 2002*, 2002.

[18] H.-H. Nagel. On the estimation of optic flow: Relations between different approaches and some new results. *Artificial Intelligence*, 33:299–324, 1987.

[19] B.A. Olshausen and D. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[20] W.A. Phillips and W. Singer. In search of common foundations for cortical processing. *Behavioral and Brain Sciences*, 20(4):657–682, 1997.

[21] S. Posch. *Perzeptives Gruppieren und Bildanalyse*. Habilitationsschrift, Universität Bielefeld, Deutscher Universitäts Verlag, 1997.

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

[23] M. Sigman, G.A. Cecchi, C.D. Gilbert, and M.O. Magnasco. On a common circle: Natural scenes aand gestalt rules. *PNAS*, 98(4):1935–1949, 2001.

[24] E.P. Simoncelli and B.A. Ohlshausen. Natural image statistics and neural representations. *Anual Reviews of Neuroscience*, 24:1193–1216, 2001.

[25] R.J. Watt and W.A. Phillips. The function of dynamic grouping in vision. *Trends in Cognitive Sciences*, 4(12):447–154, 2000.

[26] C. Zetsche and G. Krieger. Nonlinear mechanisms and higher–order statistics in biologial vision and electronic image processing: review and perspectives. *Journal of electronic imaging*, 10(1), 2001.