Statistics of Second Order Multi–Modal Feature Events and their Exploitation in Biological and Artificial Visual Systems

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

July 5, 2002

Abstract

In this work we investigate the multi-modal statistics of natural image sequences looking at the modalities orientation, color, optic flow and contrast transition. It turns out that second order interdependencies of local line detectors can be related to the Gestalt law collinearity. Furthermore we can show that statistical interdependencies increase significantly when we look not at orientation only but also at other modalities.

The occurrence of illusionary contour processing (in which the Gestalt law 'collinearity' is tightly involved) at a late stage during the development of the human visual system (see, e.g., [3]) makes it plausible that mechanisms involved in the processing of Gestalt laws 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. We discuss the usage of statistical interdependencies measured in this work within an artificial visual systems and show first results.

Keywords: Grouping, Modality Integration, Statistics of Natural Images

Statistics of Second Order Multi–Modal Feature Events and their Exploitation in Biological and Artificial Visual Systems

Abstract

In this work we investigate the multi-modal statistics of natural image sequences looking at the modalities orientation, color, optic flow and contrast transition. It turns out that second order interdependencies of local line detectors can be related to the Gestalt law collinearity. Furthermore we can show that statistical interdependencies increase significantly when we look not at orientation only but also at other modalities.

The occurrence of illusionary contour processing (in which the Gestalt law 'collinearity' is tightly involved) at a late stage during the development of the human visual system (see, e.g., [3]) makes it plausible that mechanisms involved in the processing of Gestalt laws 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. We discuss the usage of statistical interdependencies measured in this work within an artificial visual systems and show first results.

Statistics of Second Order Multi–Modal Feature Events and their Exploitation in Biological and Artificial Visual Systems

Abstract

In this work we investigate the multi-modal statistics of natural image sequences looking at the modalities orientation, color, optic flow and contrast transition. It turns out that second order interdependencies of local line detectors can be related to the Gestalt law collinearity. Furthermore we can show that statistical interdependencies increase significantly when we look not at orientation only but also at other modalities.

The occurrence of illusionary contour processing (in which the Gestalt law 'collinearity' is tightly involved) at a late stage during the development of the human visual system (see, e.g., [3]) makes it plausible that mechanisms involved in the processing of Gestalt laws 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. We discuss the usage of statistical interdependencies measured in this work within an artificial visual systems and show first results.

1 Introduction

A large amount of research has been focused on the usage of Gestalt laws in computer vision systems (overviews are given in [19, 18]). The most often applied and also the most dominant Gestalt principle in natural images is collinearity [5, 12]. Collinearity can be exploited to achieve more robust feature extraction in different domains, such as, edge detection (see, e.g., [9, 10]) or stereo estimation [4, 18]. In most applications in artificial visual systems, the relation between features, i.e., the applied Gestalt principle, has been defined heuristically based on semantic characteristics such as orientation or curvature. 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.

In the human visual system beside local orientation also other modalities such as color and optic flow are computed (see, e.g. [7]). 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., [17, 11]). This integration is manifested in the dense connectivity within brain areas in which

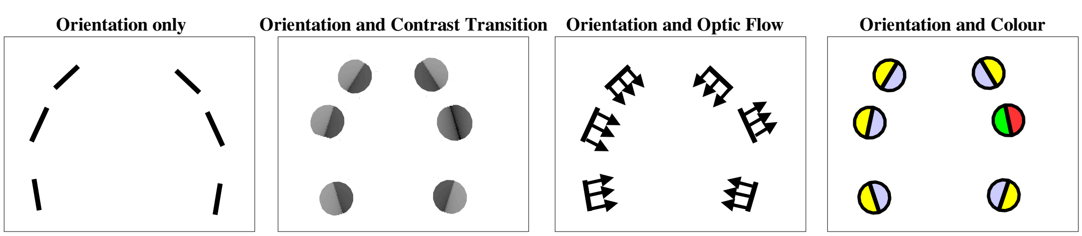

Figure 1: Grouping of visual entities becomes intensified (left triple) or weakened (right triple) by using additional modalities: Since the visual entities are not only collinear but show also similarity in an additional modality their grouping becomes more likely.

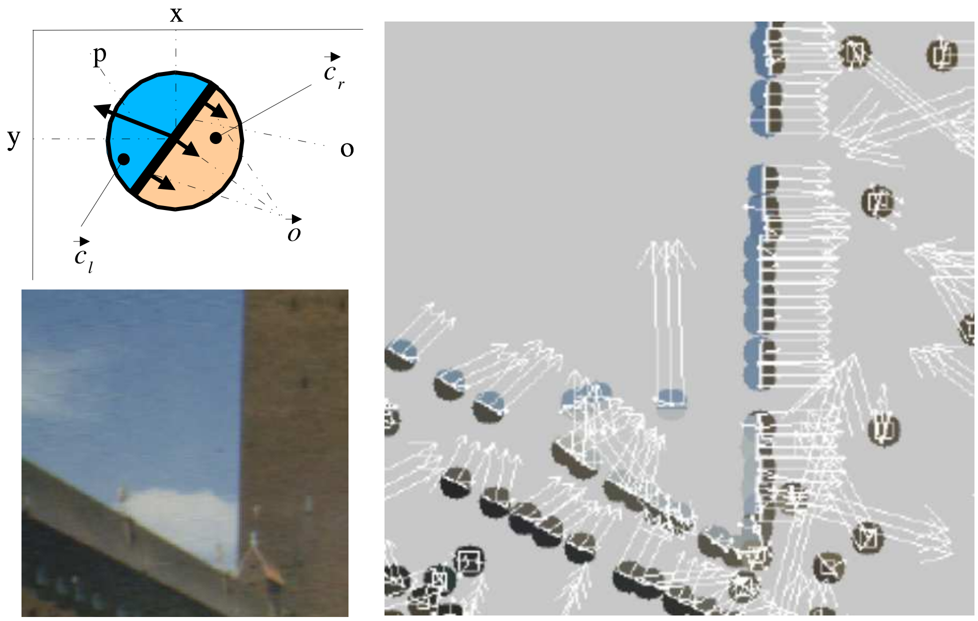

Figure 2: **Top left:** Schematic representation of a basic feature vector. **Bottom left:** Frame in an image (the frame is part of the image sequence shown in figure 3 left). **Right:** Extracted Feature vectors.

the different visual modalities are processed as well as in the large number of feedback connections from higher to lower cortical areas (see, e.g., [7]).

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 found 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 discuss how these measured interdependencies can be used within artificial visual 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 1) and 2 containing the following sub-modalities:
Orientation: We compute local orientation o (and local phase p) by the specific isotropic linear filter [6].
Contrast Transition: The contrast transition of the signal is coded in the phase p of the same filter.
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 $(c_r^l, c_g^l, c_b^l), (c_r^r, c_g^r, c_b^r)$ representing the color in RGB space on the left and right side of the edge.
Optic Flow: Local displacements (f_1, f_2) are computed by a well known optical flow technique ([15]).

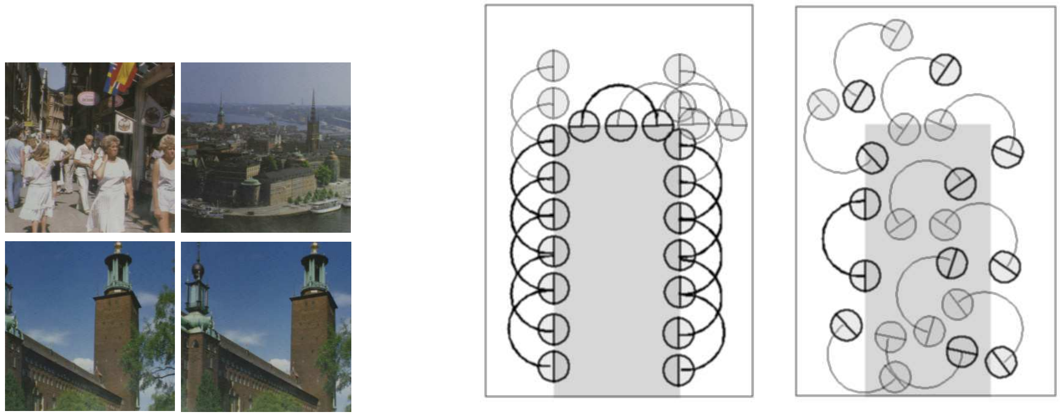

Figure 3: **Left:** Images of the data set (top) and 2 images of a sequence (bottom). Explanation of the Gestalt coefficient $G(e^1|e^2)$: We define e^2 as the occurrence of a line segment with a certain orientation (anywhere in the image). **Right:** Let the second order event e^1 be: “occurrence of 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 $G(e^1|e^2) = 0.64/0.09 = 7.1$.

All modalities are extracted from a local image patch[1]. The output is a local interpretation of the image patch by semantic properties (such as orientation and displacement) in analogy to the sparse output of a V1 column in visual cortex.

For our statistics we use 9 image sequences with a total of 42 images of size 512x512 (18 images) and 384x288 (24 images). Our data (see figure 3 left for examples) contains variations caused by object motion as well as camera motion. 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.

3 Statistical Interdependencies in Image Sequences

The Gestalt coefficient 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

[1]In our statistical measurements we only use image patches corresponding to intrinsically one-dimensional signals (see [22]) since orientation is reasonably defined for these image patches only.

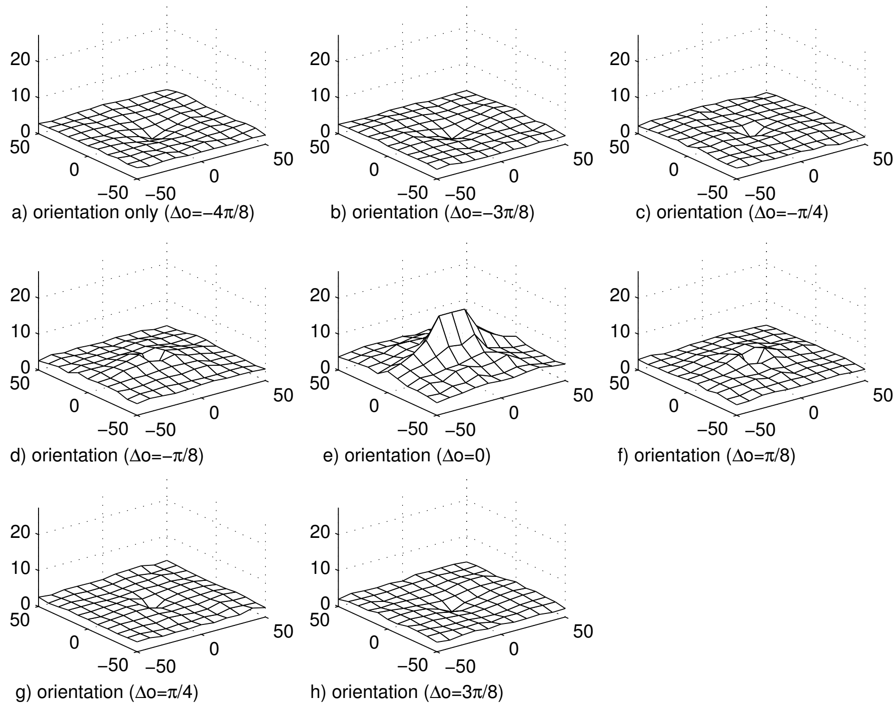

Figure 4: The Gestalt coefficient for differences in position from -50 to 50 pixel in x– and y–direction when orientation only is regarded. In a) the differecne of orientation of the line segments is $\frac{\pi}{2}$ (the line segments are ortogonal) while in e) the difference of orientation is 0, i.e., 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.

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 3. Further details can be found in [13].

3.1 Second Order Relations Statistics of Natural Images

A large amount of work has addressed the question of efficient coding of visual information and its relation to the statistics of images. Excellent overviews are given in [22, 21]. 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. [16, 2]), recently statistical investigation in the feature space of local line segments have been performed (see, e.g., [12, 5, 8]) and have addressed the representation of Gestalt principles in visual data.

Here we go one step further by investigating the second order relations of events in our *multi–modal feature space*

$$e = ((x, y), o, p, ((c^l_r, c^l_g, c^l_b), (c^r_r, c^r_g, c^r_b)), (f_1, f_2)).$$

In our measurements we collect second order events in bins defined by small patches in the

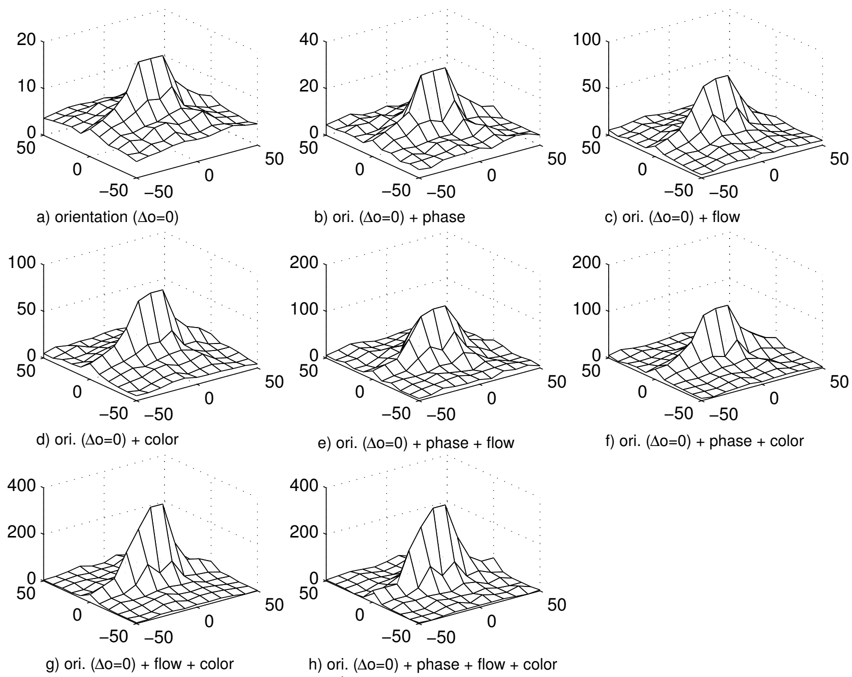

Figure 5: The Gestalt coefficient for $\Delta o = 0$ and all possible combination of modalities.

(x_1, x_2)–space and by regions in the modality–spaces defined by the metrics defined for each modality (for details see [13]). Figure 4 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 [12, 8] collinearity can be detected as significant second order relation as a ridge in the surface plot for $\Delta o = 0$ in figure 4e. Also parallelism is detectable as an offset of this surface. A Gestalt coefficient significantly above one can also be detected for small orientation differences (figure 4d,f, i.e., $\Delta o = -\frac{\pi}{8}$ and $\Delta o = \frac{\pi}{8}$) corresponding to the frequent occurrence of curved entities. 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 [12]).

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 ssimilar contrast transition (i.e., 'similar phase') when $d(p, p') < t^{p+}$. The metrics in the different modalities are precisely defined in [13]. t^{p+} is defined such that only 10% of the comparisons $d(p, p')$ in the data set are smaller than t^p. Figure 5b shows the Gestalt coefficient for the events 'similar orientation and similar contrast transition'. In figure 6 the Gestalt coefficient along the x-axes in the surface plot of figure 5 is shown. The Gestalt coefficient on the x-axes correspond to the 'collinearity ridge'. The first column 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

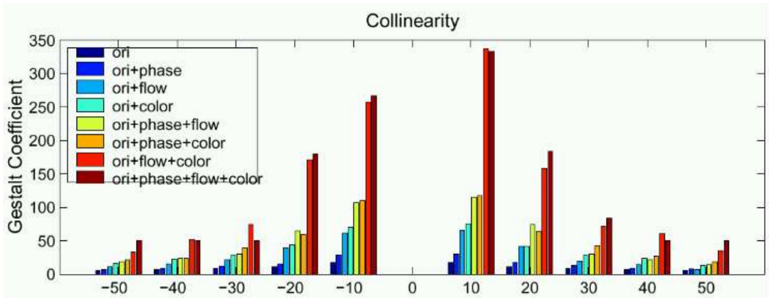

Figure 6: The Gestalt coefficient for collinear feature vectors for all combinations of modalities. For (0,0) the Gestalt coefficient is not shown, since e^1 and e^2 would be identical.

compared to the case when we look at orientation only for collinearity.

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.

Orientation and Color: Analogously, we define that two events have 'similar color structure'. The corresponding surface plot is shown in figure 5d and the slice corresponding to collinearity is shown in the fourth column in figure 6.

Orientation and Optic Flow: The corresponding surface plot is shown in figure 5c and the slice corresponding to collinearity is shown in the third column in figure 6. An even more pronounced increase of inferential power for collinearity can be detected.

Multiple additional Modalities: Figure 5 shows the surface for similar orientation, phase and optic flow (figure 5e); similar orientation, phase and color (figure 5f) and similar orientation, optic flow and color (figure 5g). The slices corresponding to collinearity are shown in the fifth to seventh columns in figure 6. 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 5h and the eighth column in figure 6 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.

4 Summary and Examples of Possible Applications

In this paper we have addressed the statistics of local oriented line segments derived from natural scenes by adding information to the line segment concerning the modalities contrast transition, color, and optic flow. We could show that statistical interdependencies in the orientation–position domain correspond to the Gestalt laws collinearity and parallelism and that they become significantly stronger when multiple modalities are taken into account. Essentially it seems that visual information bears a high degree of intrinsic redundancy. This redundancy can be used to reduce the ambiguity of local feature processing.

The results presented here provide further evidence for the assumption that despite the vagueness of low level processes stability can be achieved by *integration of information across modalities*. In addition, the attempt to model the application of Gestalt laws based on statistical measurements, as suggested recently by some researchers (see, [8, 5, 12, 20]) gets further support. Most importantly, the results derived in this paper suggest to formulate the application of Gestalt principles in a multi-modal way.

Illusionary contour processing (in which the Gestalt law 'collinearity' is tightly involved) occurs at a late stage (after approximately 6 months) during the development of the human visual system (see [3] and [14]). This late development of the above mentioned mechanisms makes it likely

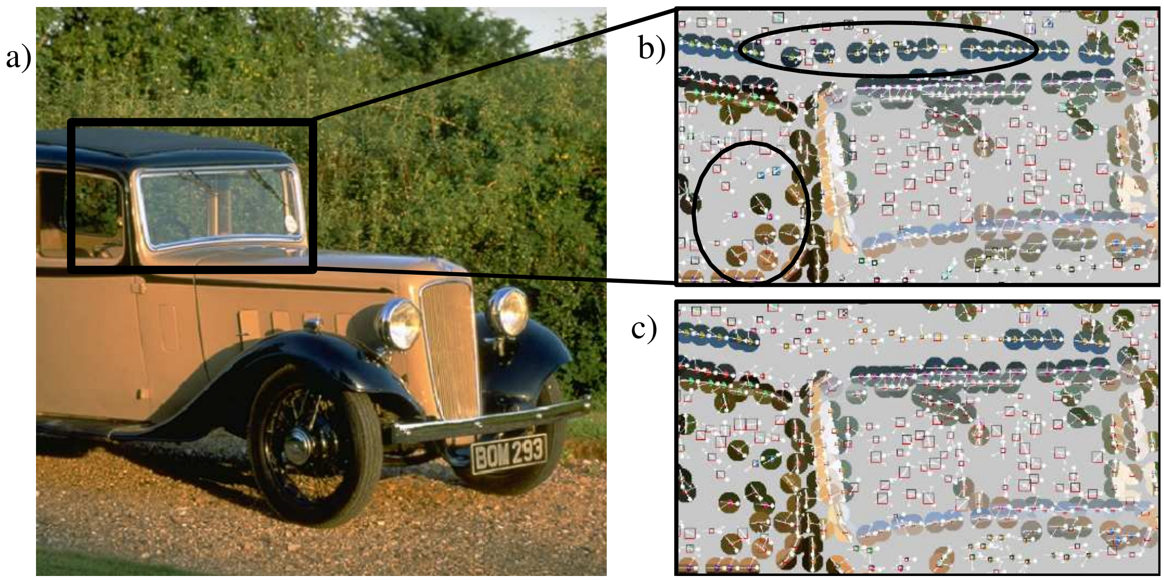

Figure 7: Left: Image of a car. Right top: Extraction of features with grouping based on the Gestalt coefficient. Right bottom: Feature extraction without grouping.

that those mechanisms depend on visual experience of the underlying structures in visual data. This also suggests a formalization of Gestalt laws in artificial systems depending on statistical measurements.

Motivated by the measurable reflectance of Gestalt principles in the statistics of natural images (as shown in this paper) and the late development of abilities in which these Gestalt principles are involved, it is our aim to replace heuristic definition of Gestalt rules by interaction schemes based on statistical measurements. 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.

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 7b and c all features above a the very same threshold are displayed with a filled circle while features below this threshold are displayed without these circles. Note the detection of the low contrast edge (figure 7b, horizontal ellipse) when applying grouping based on the Gestalt coefficient and the reduction of non meaningful features (vertical ellipse) without grouping.

The exact formalization of grouping and feature disambigution based on the statistical measurements explained here is part of our current research and will be described in a forthcoming 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] R.C.K. Chung and R. Nevatia. Use of monucular groupings and occlusion analysis in a hierarchical stereo system. *CVPR*, 1991.

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

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

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

[8] 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.

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

[10] 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.

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

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

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

[14] 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.

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

[16] 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.

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

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

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

[20] 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.

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

[22] 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.