INSTITUTE OF PHYSICS PUBLISHING NETWORK: COMPUTATION IN NEURAL SYSTEMS

Network: Comput. Neural Syst. **13** (2002) 1–23 PII: S0954-898X(02)53197-3

Multi-modal estimation of collinearity and parallelism in natural image sequences*

Norbert Krüger and Florentin Wörgötter

Institute for Computational Intelligence and Technology (INCITE), University of Stirling, Scotland FK9 4LA, UK

E-mail: norbert@cn.stir.ac.uk

Received 9 January 2001, in final form 22 August 2002
Published
Online at stacks.iop.org/Network/13/1

Abstract

In this paper we address the statistics of second-order relations of feature vectors derived from image sequences. We compute the individual vector components corresponding to the visual modalities orientation, contrast transition, optic flow, and colour by conventional low-level early vision algorithms. As a main result, we observe that collinear (or parallel) line pairs are, with very great likelihood, also associated with other identical features, for example sharing the same flow pattern, or colour or even sharing multiple feature combinations.

It is known that low level processes, such as edge detection, optic flow estimation and stereo are ambiguous. Our results provide support for the assumption that the ambiguity of low level processes can be substantially reduced by integrating information across visual modalities. Furthermore, the attempt to model the application of gestalt laws in computer vision systems based on statistical measurements, as suggested recently by some researchers (Elder H and Goldberg R M 1998 *Perception Suppl.* **27**; Geisler W S, Perry J S, Super B J and Gallogly D P 2002 *Vis. Res.* **41** 711–24; Krüger N 1998 *Neural Process. Lett.* **8** 117–29; Sigman M, Cecchi G A, Gilbert C D and Magnasco M O 2001 *Proc. Natl Acad. Sci. USA* **98** 1935–49), gets further support and the results in this paper suggest formulation of gestalt principles in artificial vision systems in a multi-modal way.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Substantial research has been focused on the usage of gestalt laws in computer vision systems (excellent overviews are given in [9, 62]). The most often applied gestalt principle in artificial visual systems and also the most dominant gestalt principle in the 2D projection of natural scenes is collinearity [14, 21, 45, 66]. Collinearity can be exploited to achieve more robust

See endnote 1

See endnote 2

* This work has, to a large part, been performed during Norbert Krüger's stay in the Cognitive System Group at the University of Kiel, Germany.

0954-898X/02/000001+23$30.00

Processing CRC data

NET/net153197-xsl/PAP File name NE .TEX First page
Printed 19/9/2002 Date req. Last page
Issue no. Total pages

Focal Image
(Ed: STUART)

feature extraction in different domains, such as edge detection (see, e.g., [25, 29, 36]) or stereo estimation [11, 58].

In most applications of 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 (e.g. two line segments are defined to be collinear when they lie on a contour with slowly changing curvature [77]). Mostly, explicit models of feature interaction have been applied, connected with the introduction of parameters to be estimated beforehand, a problem recognized as awkward in computer vision. Recently, Geisler *et al* [21] introduced the idea to overcome heuristic and explicit models by relating feature interaction to the statistics of natural images. The feasibility of this approach receives strong support from the *measurable interdependencies* of features in visual scenes as performed here and in some recent work [14, 21, 45, 66]. The necessity of an adaptive component in the formalization of statistical interdependencies is also supported by the fact that the human visual system generates these patterns of feature interaction by visual experience: developmental psychology shows strong evidence that visual experience plays an important role in achieving the ability to use these interdependencies in visual processing (e.g. the effect of illusionary contours appears after 5 months [7, 38]).

In the human visual system, apart from local orientation other modalities[1] such as colour and optic flow are also computed (see, e.g., [20]). All these low level processes face the problem of a high degree of vagueness and uncertainty [1]. This arises from a couple of factors. Some of them are associated with image acquisition and interpretation: owing to noise in the acquisition process along with the limited resolution of cameras, only inaccurate estimates of semantic information (e.g. orientation) are possible. Furthermore, illumination variation heavily influences the measured grey level values and is hard to model analytically [33]. Information extracted across image frames, e.g. in stereo and optic flow estimation, faces (in addition to the above-mentioned problems) the correspondence and aperture problem which interfere in a fundamental and especially difficult way (see, e.g., [2, 39]). However, the human visual system acquires visual representations which allow actions with high precision and certainty within the 3D world under rather uncontrolled conditions. It can achieve the required certainty and completeness by integrating visual information across modalities (see, e.g., [30, 57]). The power of modality fusion arises from the *huge intrinsic relations* given by regularities within and across visual modalities. The essential need for integrating visual information, in addition to optimizing single visual modalities to design efficient artificial visual systems, has also been recognized in the vision community after a long period of work on improving single modalities [1].

Gestalt principles are also affected by multiple visual modalities. For example, figure 1 shows how grouping based on the gestalt law collinearity can be intensified by the different modalities contrast transition, optic flow and colour. This paper addresses the statistics of natural images in these modalities. As a main result we report that statistical interdependencies in visual scenes become significantly stronger when multiple visual modalities are taken into account. This result gives further evidence for the assumption that, despite the vagueness of low level processes, stability can be achieved by integration of information across modalities. Furthermore, the attempt to model the application of gestalt laws based on statistical measurements [14, 21, 45, 66] gets further support. Finally, the results in this paper suggest the formulation of applications of gestalt principles in a multi-modal way.

[1] In the following we use the term 'modality' to refer to visual modalities such as colour or stereo and we do not refer to other sensorial modalities such as touch or sound.

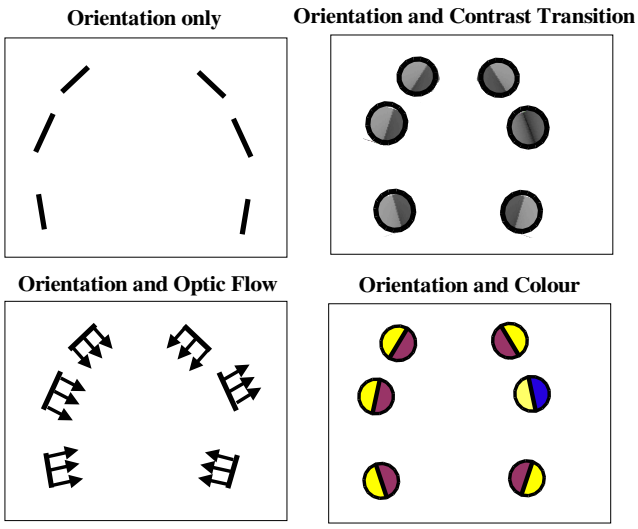

Figure 1. Grouping of entities becomes intensified (left triple) or weakened (right triple) by using additional visual modalities.

2. Literature on gestalt principles and natural images and contribution of this work

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 [67, 73, 78]. 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., [6, 28, 56]), recently statistical investigations in more complex feature spaces have been performed (see, e.g., [14, 21, 45, 66]) and have addressed the reflection of gestalt principles in these feature spaces.

Collinearity and parallelism are two examples of gestalt laws (see, e.g., [42, 71]). Gestalt laws reflect regularities of objects and object constellations caused by physical and biological factors in the 3D world (such as gravity, growth and erosion) and projections thereof. Collinearity and parallelism describe *probabilistic* feature relations. More specifically, the occurrence of a line segment in visual data has a distinct impact on the likelihood of the occurrence of a specific line segment at a different position. These probabilistic relations can be *measured* from natural images (see [14, 21, 45, 66]). In [45], it has also been shown that the distinctness of second-order dependencies between local oriented entities depends significantly on their processing: while in the space of Gabor wavelet responses the statistical dependencies were barely detectable, a post-processing which interprets Gabor wavelets as line segment detectors[2] made collinearity and parallelism distinctly visible in the second-order statistics of natural images (see figure 2 and [45]).

Decades ago, Brunswick and Kamiya [10] first 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. The strong prevalence of collinearity in natural images has been investigated first by [45] and [14]. These results have been confirmed and extended by [21, 66]. In addition to [45] and [14], in [66] a co-circularity rule has been established as a generalization of collinearity, which says that, after straight lines, circular structures form the most common second-order relations of line segments in

[2] This post-processing played an important role in the object recognition system ORASSYLL [47].

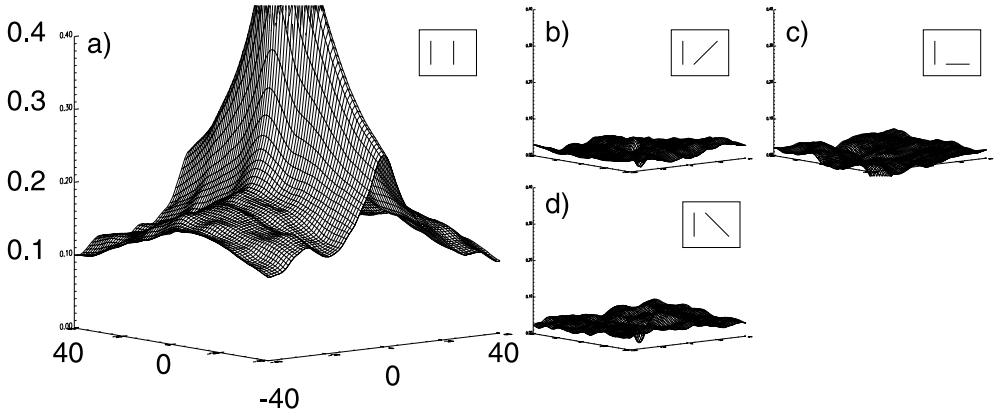

Figure 2. The cross-correlation of pairs of post-processed Gabor wavelet responses (a)–(d) of four orientations on a large set of natural images: (a) horizontal–horizontal, (b) horizontal–diagonal, (c) horizontal–vertical, (d) horizontal–diagonal. The x and y axes represent the separation of the kernels (labeling of all axes for (a)–(d) is the same as in (a) and the z axis represents the correlation. In (a) parallelism and collinearity are clearly visible: collinearity is detectable as a ridge in the first diagram and parallelism appears as a global property expressed in the flat part of the surface in the first diagram clearly above the surfaces corresponding to non-parallel orientations. In contrast, the correlation to non-similar orientations is low (b)–(d) (more detailed results can be found in [45]).

natural images. While [45, 66] have investigated the second-order relation of line segment responses without considering whether the two line segments belong to the same 3D contour or not (called 'absolute co-occurrence' in [21]), [14, 21] have investigated conditional densities which take this contour coincidence into account (so-called 'Bayesian co-occurrence'). This distinction is especially important since only in the case of a collinearity event caused by a 3D contour is the grouping of entities justified. However, [21] have shown that in both cases the second-order statistics are similar, which indicates that collinearity is mostly determined by projections of 'real' 3D contours. In addition to collinearity and parallelism, Elder and Goldberg [14] show that other gestalt cues (proximity and luminance similarity) can also be related to the statistics of natural images.

This work is concerned with the statistical interdependencies of *specific position/orientation* relations of local oriented visual filters. It has been shown by [65] that there is a *general law* holding for linear filters in different sensorial domains that states that the standard deviation of one linear filter response scales linearly with the amplitude of another neighboured linear filter response. They show that this is a property of linear filters applied to natural signals in general, independent of the spatial relation or their relative orientation differences. However, they also note that the strength of this effect depends on the relative position and orientation. Here and in [46] we investigate such *quantitative differences* explicitly and we relate them to the gestalt principles of collinearity and parallelism.

In the work presented here we address the multi-modal statistics of natural images. We start from a feature space (see also figure 1) which is motivated by feature processing in V1 (see, e.g., [20]) and is described in detail in section 4.2:

Orientation

We compute local orientation (and local phase) by the isotropic linear filter [18].

Contrast transition

The contrast transition of the signal is coded in the phase of the applied filter.

Optic flow

Local displacements are computed by a well known optical flow technique [55].

Colour

Colour 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).

All visual modalities are extracted from a local image patch[3], resembling columns in V1 responsible for a certain retina patch. There is ample evidence that these modalities are processed in early stages of visual processing (see, e.g., [20, 32, 35]). The feature vector represents a local interpretation of the image patch by semantic properties (such as orientation and displacement) similar to the sparse output of a V1 column (see section 4.2).

One main contribution of this paper is to show that statistical interdependencies of collinear local line segments can be increased significantly by making use of the additional modalities colour, contrast transition and optic flow. We evaluate the inferential power of each modality separately (measured by the so-called *gestalt coefficient*, see section 3 and [45]) as well as the inferential power of joint modalities. In our multi-modal feature space, we show that large redundancies (much higher than when orientation only is considered) exist which become even stronger when multiple visual modalities are taken into account at the same time. These redundancies correspond to the gestalt principles collinearity and parallelism and reflect a strong inferential power between our visual entities. We suggest that these redundancies can be used by biological and artificial visual systems to overcome the uncertainty which is inherent in the local processing of features. In this sense, the statistical investigations introduced here, combined with the idea of relating statistical interdependencies to feature interaction rules [14, 21, 45, 66], can be seen as a preparatory work for the design of more sophisticated and efficient feature interaction across multiple visual modalities (see also section 6.2).

Another interesting result concerns the statistics of an intrinsically 1D structure: using specific pre-processing which performs a split of an intrinsically 1D signal into geometrical and structural information [18] we can characterize the distribution of intrinsically 1D signals in natural images. It turns out that edge structures are much more dominant than line structures.

The paper is organized as follows: in section 3 we introduce the measure for interdependence used in our statistical investigation. In section 4 we describe our feature space as well as the data used in our simulations. Then, in section 5 we describe the results of our simulations. In section 6 we discuss our results in the context of the realization of gestalt principles in biological systems and their formalization within artificial systems.

3. Measuring statistical interdependencies

We measure statistical interdependencies between events by a mathematical term that we call the 'gestalt coefficient'. 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:

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

[3] In our statistical simulations we only use image patches corresponding to intrinsically one-dimensional signals since orientation is reasonably defined for these image patches only. Intrinsically one-dimensional signals are constant in one orientation (see, e.g., [24, 78]).

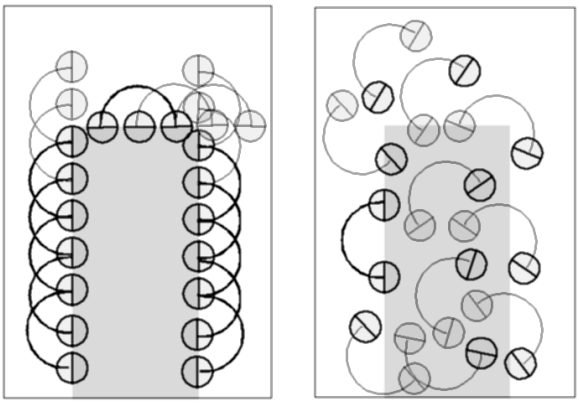

Figure 3. Explanation of the gestalt coefficient $G(e^1|e^2)$: we define e^2 as the occurrence of a line segment with a certain orientation (anywhere in the image). Let the second-order event e^1 be: 'occurrence of a line segment two units forward from an existing line segment e^2 with the same orientation as e^2'. Left: computation of $P(e^1|e^2)$. All potential occurrences of events e^1 in the image are shown. Bold arcs represent real occurrences of the specific second-order relations e^1 whereas grey arcs 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 a line segment two units forward from an existing line segment e^2 with the same orientation then e^2' is evaluated. (In our simulations we used more than 500 000 samples for the estimation of $P(e^1)$.) Only in 1 of 11 possible cases does 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 $G(e^1|e^2) = 0.64/0.09 = 7.1$.

For the modelling of feature interaction a high gestalt coefficient indicates an increase in the likelihood of the event e^1 depending on other events e^2: since

$$P(e^1|e^2) = G(e^1, e^2)P(e^1)$$

the gestalt coefficient says how the likelihood of the event e^1 is modified by occurrence of the event e^2. 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.

In our case we are interested in the events e^2 'occurrence of a line segment with a certain orientation (anywhere in the image)' and e^1 'occurrence of a line segment two units forward from an existing line segment e^2 with the same orientation as e^2' (the gestalt coefficient for these events is illustrated in figure 3). Since we are only interested in the relative position and orientation relation of the two events we can also express $G(e^1, e^2)$ by

$$G(\Delta x_1, \Delta x_2, \Delta o) = G(e^1, e^2).$$

Note that the gestalt coefficient is related to the correlation of two random variables (see [45]) as well as to the mutual information contained in two symbols used in information theory (see, e.g., [27]). The mutual information is the logarithm of the gestalt coefficient.

4. Visual modalities and data

In our investigation we make use of the four visual modalities: orientation o, contrast transition (expressed by the phase p), colour on the left and right side of the edge ($\vec{c}_l$, $\vec{c}_r$) and optic flow $\vec{o}$.

Figure 4. Top: images of the data set. Bottom: three images from a sequence.

Together with the position $\vec{x}$, these modalities create a feature vector

$$e = (\vec{x}, o, p, (\vec{c}_l, \vec{c}_r), \vec{o})$$

which describes the local area of an image patch (see figure 5, top left). In this section we describe the data used for our statistics and the preprocessing of this feature vector.

4.1. Data

For our statistics we use 95 images from 20 image sequences of size 512 × 512 (11 images), 384 × 288 (36 images), and 359 × 273 (48 images). Our data set contains indoor as well as outdoor scenes (see figure 4)[4]. The results of our statistical investigation were similar for the indoor and outdoor sequences, i.e. they were not dominated by a specific subset. The image sequences contain variations caused by object motion as well as camera motion. There were more than 30 000 feature vectors detected in the data set (approximately 20 000 from the outdoor images).

4.2. Feature processing

Edge detection and orientation estimation is based on the isotropic linear filter (called a monogenic signal [18]) and on phase congruence over neighbouring frequency bands (see e.g., [17, 44]). The monogenic signal performs a *split of identity* [18]: it orthogonally divides an intrinsically one-dimensional bandpass filtered signal into energetic information (indicating the likelihood of the presence of a structure), its geometric information (orientation) and its contrast transition (called 'structure' in [17]). We look for energy maxima in the position–orientation space $(\vec{x}, o)$. We use hexagonally arranged patches with a diameter of 9 pixels. To avoid the occurrence of very close line segments produced by the same image structure we demand that line segments have a certain minimal distance.

The variance of orientation in an image patch (computed from pixel positions of high energy) is indicated as a rectangle in the displays of feature vectors in figure 5, right. We use

[4] The outdoor scenes can be downloaded from the internet via http://www.inrialpes.fr/movi/pub/Images/and http://sampl.eng.ohio-state.edu/∼sampl/database.htm.

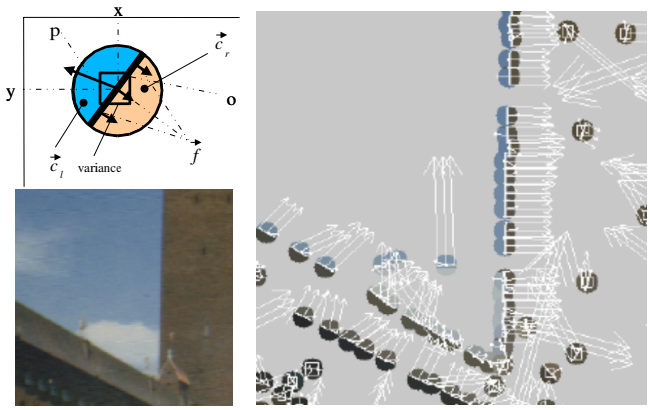

Figure 5. Top left: schematic representation of a basic feature vector. Bottom left: frame in an image. Right: extracted feature vectors.

it as a measure for the intrinsic dimensionality. In our simulations we only use features for which the variance of orientation within a small patch is below a certain threshold t^{iD} (in our case $t^{iD} = 0.5$). The distribution of orientations is not isotropic (see figure 7, left). We can see maxima for horizontal and vertical orientation.

Orientation. As for all other modalities which we investigate here, we can define a metric $d(o, o')$, i.e. we can speak of similarity and dissimilarity in the modality orientation. This metric allows pooling of similar events. Note that metric organization is also a well established design principle in biological visual systems (see, e.g., [40]). The metric for orientation and all other modalities is defined in appendix A.

Contrast transition. Contrast transition is coded in the phase at a local maximum in the (x, y, o) feature space [44]. It refers to the kind of intrinsically one-dimensional grey level structure existent at the local image patch (as a dark/bright edge, or 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 (see figure 6). Therefore, it allows coding different kinds of edge-like structures by one parameter. A metric for the phase is defined in appendix A. This metric plays a crucial role in a stereo algorithm described in [49].

The distribution of phases in our data is shown in figure 7 (middle). The peaks at $p = \pi/2$ and $-\pi/2$ show that edges (i.e. intrinsic one-dimensional signals with odd symmetry) are the dominant one-dimensional structure in natural images while line structures (i.e. intrinsic one-dimensional signals with even symmetry) are less dominant. Our model for an intrinsically one-dimensional signal patch (see figure 5) therefore describes edges[5,6].

In this paper we will show that the inferential power of oriented entities in images becomes heavily increased by making use of phase (i.e. by looking at contrast transition) *in addition* to orientation. This result is in accordance with the study of Elder and Goldberg [14] who have

[5] Although there are significantly more edge-like structures than line-like structures in natural images it may also make sense to introduce an extra line model to describe intrinsically one-dimensional image patches with phase close to 0 or π.

[6] The phase as an additional feature allows us to take the grey level information in an image patch, in addition to orientation, into account in a very compact way. The argument between purely geometric and appearance-based representations (see, e.g., [54]) becomes dissolved by integrating both descriptions within one representation. This becomes possible by the split of identity of the monogenetic signal [18] and the interpretation of phase as characterization of the contrast transition [18, 44].

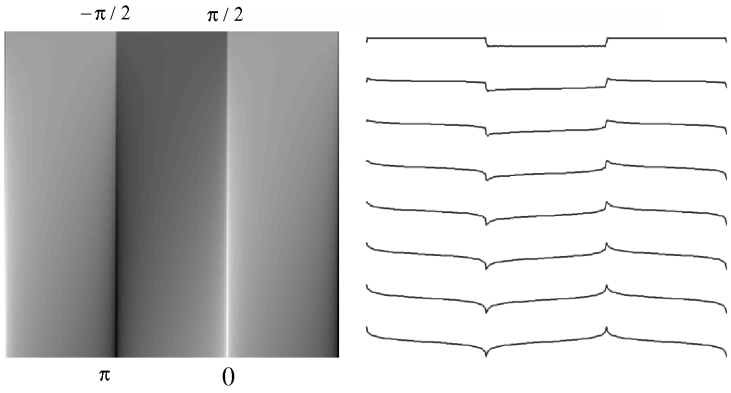

Figure 6. Left: variation of contrast transition according to phase variation. Note that a phase of π codes a dark line on a bright background, a phase of $-\pi/2$ coded a bright/dark edge, a phase of 0 codes a bright line on a dark background while a phase of $\pi/2$ codes a dark/bright edge. As can be seen the continuum between these case is also coded by the phase. Right: luminance profiles corresponding to the image on the left.

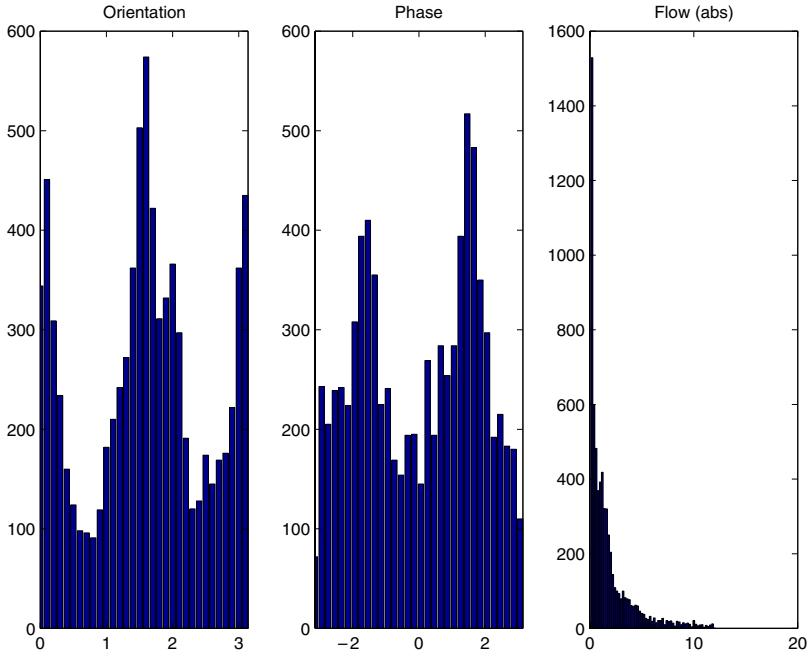

Figure 7. Distribution of orientation (left), phase (middle) and magnitude of optic flow (right) in the data set.

also investigated luminance structure but using a different kind of description based on the differences of luminance at both sides of the edge.

Colour. To integrate the modality colour at intrinsically one-dimensional image structures we perform a Gaussian integration in the RGB colour space over the left and right part ('left' and 'right' defined by the associated line segment) of the image patch (see figure 5). Since

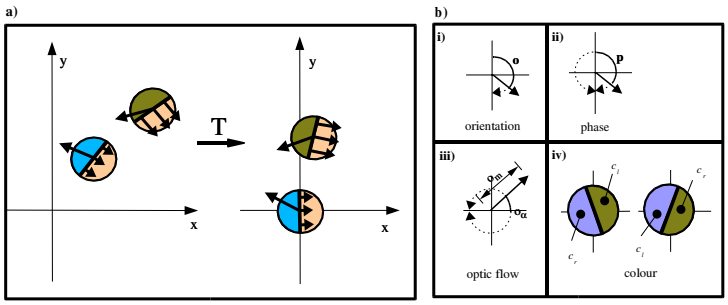

Figure 8. (a) Transformation of the coordinate system such that e^2 has position $(0, 0)$ and orientation 0. (b) The coordinate systems for the different modalities.

the distribution of phases indicates the dominance of edges (figure 7), this kind of integration corresponds to the most likely model of intrinsically one-dimensional structures. We get two vectors $\vec{c}_l = (c_r^l, c_g^l, c_b^l)$ and $\vec{c}_r = (c_r^r, c_g^r, c_b^r)$, representing the red, green and blue values of the left and right sides of the edge. A metric for colour is defined in appendix A.

Optic flow. Optic flow $\vec{o}$ is computed by a differential based method [55] which gives especially good results at edges. We transfer the optic flow vector $(\vec{o}_x, \vec{o}_y)$ to a representation of magnitude and angle $(\vec{o}_m, \vec{o}_\alpha)$ (see figure 8(b),(iii)). Figure 7 shows the distribution of optic flow magnitudes. Since objects and/or camera move slowly, small displacements dominate the data set.

5. Second-order relations statistics of natural images

In this subsection we investigate the second-order relations of events in our feature space

$$(\vec{x}, o, p, (\vec{c}_l, \vec{c}_r), \vec{o}) = ((x_1, x_2), o, p, ((c_r^l, c_g^l, c_b^l), (c_r^r, c_g^r, c_b^r)), (o_m, o_\alpha)).$$

5.1. Evaluation of known interdependencies in the features orientation and space

The distribution of orientations of the extracted multi-modal feature vectors is non-isotropic (see figure 7 and [45]) with a significantly higher density for vertical and horizontal orientation than for diagonal orientation. For our second-order statistics this plays no role. 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 and is positioned at the origin (see figure 8(a)). The exact transformation is given in appendix B. In the following we assume this transformation always applied.

In our simulations we use a discretization of the position–orientation space in bins of size 10×10 and $\frac{1}{8}\pi$. Figure 9 shows the gestalt coefficient when we use only orientation as a visual modality.

As in figure 2 and as already shown in [21, 45] collinearity can be detected as a significant second-order relation by a ridge in the surface plot for $\Delta o = 0$ in figure 9(e). Also parallelism is detectable as a slight offset of this surface. A gestalt coefficient significantly above one can also be detected for small orientation differences (figures 9(d), (f), i.e. $\Delta o = -\frac{\pi}{8}$ and $\frac{\pi}{8}$). The general shape of surfaces is similar in all the following simulations concerned with additional visual modalities: we find a ridge corresponding to collinearity and an offset corresponding to

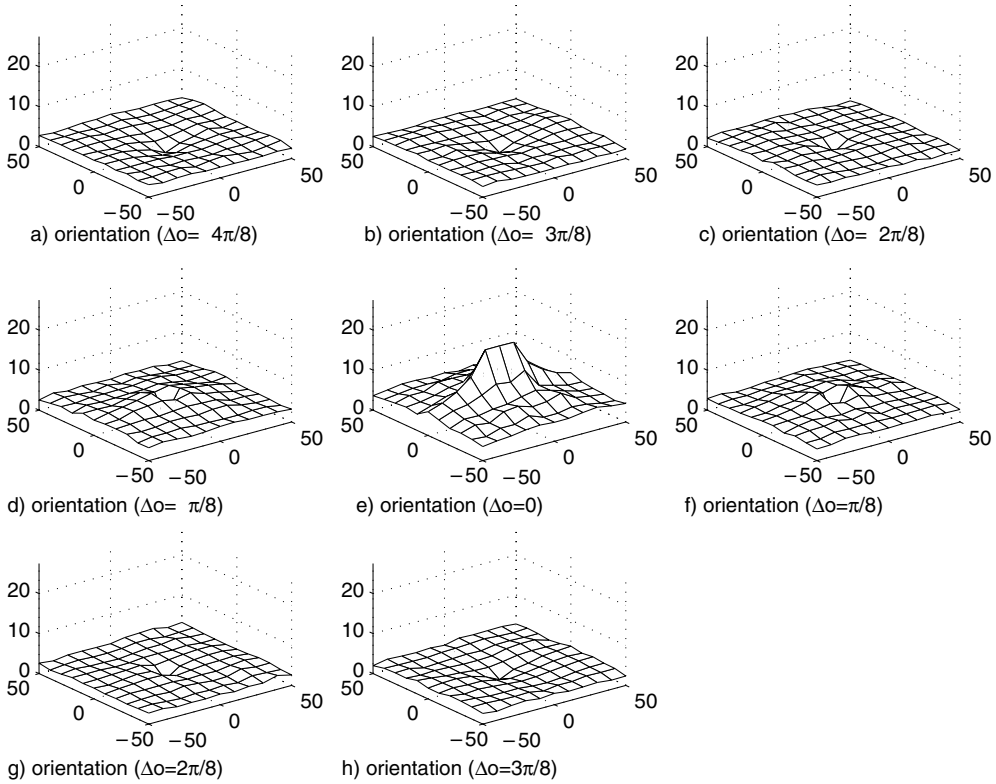

Figure 9. The gestalt coefficient for differences in position from −50 to 50 pixel in the x and y direction when orientation only is regarded. Note that the gestalt coefficient for position (0,0) and $\Delta o = 0$ is set to the maximum of the surface for better display. The gestalt coefficient is not interesting at this position, since e^1 and e^2 are identical.

parallelism and a gestalt coefficient close to one for all larger orientation differences. Therefore, in the following we will only consider the surface plots for equal orientation $\Delta o = 0$.

5.2. Pronounced interdependencies by using additional visual modalities

Now we can look at the gestalt coefficient when we also take the modalities contrast transition, optic flow and colour into account. We are interested in whether and how a certain property of an oriented entity e^1 in one of the additional visual modalities allows for predictions for the other oriented entities e^2 *in their associated properties*.

Orientation and contrast transition. We define that two events $((x_1, x_2), o, p)$ and $((x_1', x_2'), o', p')$ undergo a similar 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 (in our case $t^p = 0.13$). We now compute the gestalt coefficient $G(e^1, e^2)$ (or $G(\Delta x_1, \Delta x_2, \Delta o)$) for line segments with similar contrast transition. More specifically, assume we have extracted an oriented edge e^1 with associated contrast transition p at a certain position (event e^1), so the event e^2 is the occurrence of a line segment with similar associated contrast transition at a certain relative position $(\Delta x_1, \Delta x_2)$ and orientation (Δo). The exact definition of our computation is given in appendix C.

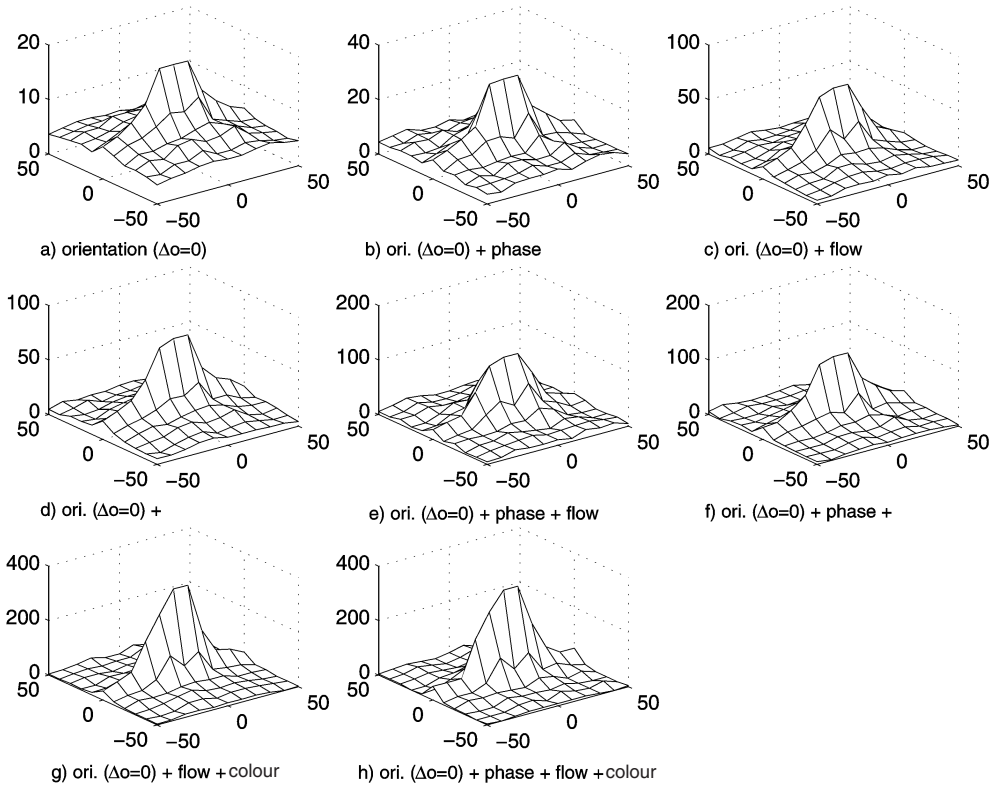

Figure 10. The gestalt coefficient for $\Delta o = 0$ and all possible combinations of visual modalities. Notice the change in scale.

Figure 10(a) replots figure 9(e). In comparison figure 10(b) shows the gestalt coefficient for the events 'similar orientation and similar phase'. In figure 11 the gestalt coefficient along the x and y axes in the surface plots of figure 10 are shown. The gestalt coefficient on the x axes correspond to the 'collinearity' ridge and the gestalt coefficient on the y axes corresponds to parallelism in the orthogonal direction to the line segment. The leftmost bars represent the gestalt coefficient when we look at similar orientation only (i.e. figure 10(a)), while the bars on the right represent the gestalt coefficient when we look at similar orientation and similar phase (i.e. figure 10(b)). We see a significant increase of the gestalt coefficient compared to the case when we look at orientation only for collinearity and only a moderate increase (or even decrease) for parallelism. This indicates that the visual modality orientation and contrast transition can be used to predict entities that are collinear (and parallel). More precisely, once we have extracted an oriented edge with an associated 'contrast transition' we are able to predict spatially distinct collinear (and parallel) edges with likelihoods expressed by the gestalt coefficient. In addition, we can predict that it is also likely that they are not only collinear but also share similar 'contrast transition'.

Orientation and optic flow. Analogously, we say that two events have 'similar flow' when $d(\vec{o}, \vec{o}') < t^f$ ($t^f = 0.17$, defined analogously to t^p and t^c). The corresponding surface plot is shown in figure 10(c) and the two slices corresponding to collinearity and parallelism are shown as the third bar in figure 11. An even more pronounced increase of inferential power for collinearity can be detected.

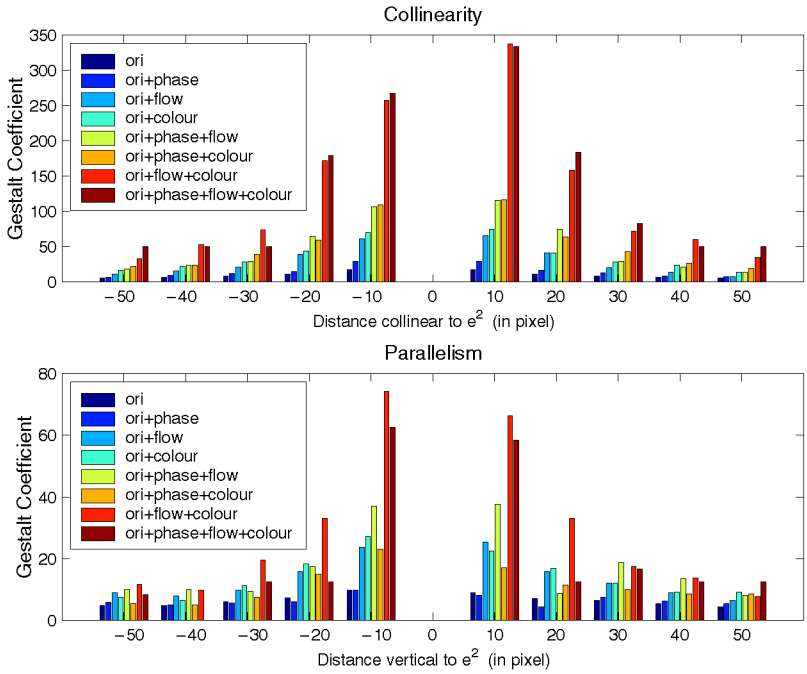

Figure 11. The gestalt coefficient for collinear and parallel feature vectors for all combinations of visual modalities. For (0, 0) the gestalt coefficient is not shown, since e^1 and e^2 would be identical.

Here we would like to remark that the significance of the inferential power of optic flow is context dependent. In all our pictures we have object or ego-motion. In the case of only little motion a smaller statistical interdependency is to be expected. In complex systems this may lead to situation-dependent statistics of interdependencies and their application.

Orientation and colour. Analogously, we say that two events have 'similar colour structure' when $d(c, c') < t^c$ ($t^c = 0.13$ again is defined such that only 10% of the comparisons $d(c, c')$ in the data set are smaller than t^p). The corresponding surface plot is shown in figure 10(d) and the two slices corresponding to collinearity and parallelism are shown as the fourth bar in figure 11.

Multiple additional visual modalities. Figure 10 shows the surface for similar orientation, phase and optic flow (figure 10(e)); similar orientation, phase and colour (figure 10(f)) and similar orientation, optic flow and colour (figure 10(g)). The slices corresponding to collinearity and parallelism are shown in the fifth to seventh bar in figure 11. We can see that the gestalt coefficient for collinear line segments increases significantly, most distinctly for the combination of optic flow and colour (seventh column). Finally we can look at the gestalt coefficient when we take all three modalities into account. Figure 10(h) and the eighth column in figure 11 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.

Figure 12 shows the values for the approximation of $P(e^1|e^2)$ and $P(e^1)$ in the case of 'orientation only' and 'orientation and colour'. Note that, for 'orientation and colour', both numerator and denominator decrease compared to 'orientation only', but the decrease in the

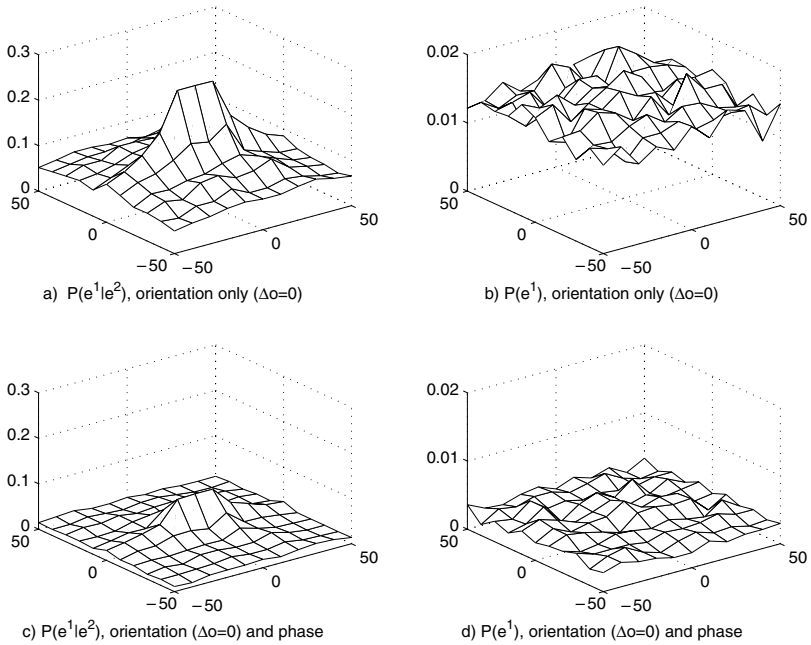

Figure 12. Top: the numerator $P(e^1|e^2)$ (left) and the denominator $P(e^1)$ (right) when we look at orientation only. Bottom: analogous to the top, but regarding orientation and phase.

denominator is larger. Note also that the estimate of the denominator $P(e^1)$ is independent of the position (see figure 12, left). Therefore we can average over the whole surface to stabilize the estimate.

6. Conclusion and discussion

In this paper we have addressed the statistics of local oriented line segments derived from natural scenes by adding information on contrast transition, colour and optic flow. We showed that statistical interdependencies in the orientation-space domain corresponding to collinearity and parallelism become significantly stronger when multiple visual modalities are taken into account. Essentially it seems that visual information bears some degree of intrinsic redundancy. Our results raise the following questions, which will be discussed now.

(1) What is the general use of such redundancies?
(2) How can these results be applied in biological and artificial visual systems?
(3) What does the preservation of redundancies imply for the understanding of visual processes in the brain?

6.1. Redundancies can be used to reduce the ambiguity of local feature processing

Vision, although widely accepted as the most powerful sensor modality, faces the problem of a high degree of vagueness and uncertainty in its low level processes such as edge detection, optic flow analysis and stereo estimation. This arises from a number of factors. Some of them are associated with image acquisition and interpretation: owing to noise in the acquisition process along with the limited resolution of cameras, only rough estimates of

semantic information (e.g. orientation) are possible. The severity of these problems even increases for higher semantic information, such as curvature (see, e.g., [5, 34]) or junction detection and interpretation (see, e.g., [26, 61]). Furthermore, illumination variation heavily influences the measured grey level values and is hard to model analytically (see, e.g., [33]). Extracting information across image frames, e.g. in stereo and optic flow estimation, faces (in addition to the above-mentioned problems) the correspondence and aperture problem which interfere in a fundamental and especially awkward way (see, e.g., [2, 39]). Furthermore, visual information is essentially incomplete since it is difficult to directly extract meaningful information at unstructured image regions because they are dominated by noise. As a result, information extracted by *local* operators is *necessarily* ambiguous. However, *by integrating information across visual modalities* (see, e.g., [1, 20, 30]), the human visual system acquires visual representations that allow for actions with high precision and certainty in most natural environments. The essential need for fusion of visual modalities, beside their improvement as isolated methods, has also been recognized by the computer vision community during the last 10 years (see, e.g., [1, 12]).

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 [14, 21, 45, 66]) gets further support. Most importantly, the results derived in this paper suggest the formulation of the application of gestalt principles in a multi-modal way (see section 6.2) since the use of additional modalities increases the statistical interdependency between visual entities. However, by making use of additional visual modalities we face the problem that the likelihood of events becomes smaller, e.g. there occur many more vertical edges in an image than blue/green vertical edges with a certain optic flow vector associated. That means the events become more meaningful and predictive but also rarer. It is known that in the human visual system visual modalities beyond orientation are computed and the effect on grouping can be demonstrated (see, e.g., [50]). Concerning the costs of coding of these rarer events, it is not necessary to code the occurrence of all multiple modality events at all times but there do exist mechanisms that allow for a *dynamic coding of feature events* such as binding and dynamic grouping [69, 70].

See endnote 3

The power of modality fusion arises from the huge number of intrinsic relations within visual scenes. The relations investigated here are concerned with statistical regularities in the sense that they allow probabilistic predictions: the occurrence of a local oriented entity with specific semantic structure makes the occurrence of another local oriented entity with certain semantic attributes at a different position *more likely*. Here we want to stress that another important regularity in visual data, with quite distinct properties compared to collinearity and parallelism, is motion, most importantly rigid body motion (RBM) (see, e.g., [16, 31]). Knowing the RBM between two frames, deterministic predictions between frames can be made (see, e.g., [15, 64]): the occurrence of an event in the first frame makes the occurrence of a certain event in the second frame mandatory (except in the special case of occlusion). While RBM leads to deterministic predictions which can also be used to stabilize feature extraction (see, e.g., [41, 48, 76]) and statistical regularities only allow probabilistic predictions we think that in a complete system both regularities have to be integrated.

6.2. Measured statistical interdependencies can be used to develop grouping algorithms

The measured multi-modal interdependencies can be used for the formalization of gestalt principles in artificial systems in a probabilistic framework (see, e.g., [13, 21]). The claim is

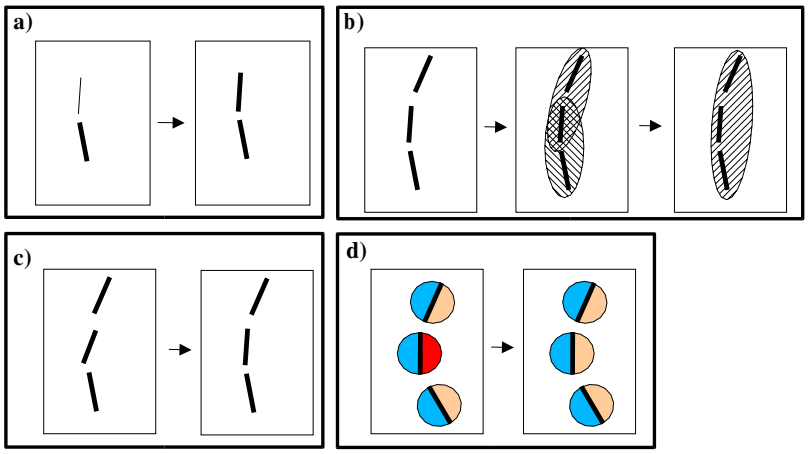

Figure 13. Schematic description of different kind of predictions. (a) Existence: the low likelihood for the existence of the entity e^1 (indicated by a thin line) increases because of the better consolidated existence of e^2 (indicated as a bold line) and the high gestalt coefficient $G(e^1, e^2)$ for 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 correction: by orientation correction (c) and colour correction (d) across entities assembled within the group a smoother variation can be achieved.

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, for example, to stabilize the system's confidence for the existence of e^1 in the 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 13(a)).

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 the 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 13(b), first arrow). As already suggested in [21], 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 then e^1 and e^3 belong to a common group as well (see figure 13(b), second arrow). Note that in this approach grouping does not require an explicitly defined higher feature constellation but is performed *dynamically*, i.e. features become assembled according to the current input [70].

Feature disambiguation. In the 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 figures 13(c) and (d)). For different visual modalities such interactions have to be formalized. Since metrics can be defined for all visual modalities (see section 4.2

and appendix A) 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 colour 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 specifics of the visual modalities also 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 will be part of future research.

In the formalization of gestalt principles suggested above an already measured gestalt coefficient is assumed. In a biological system such a separation between measuring and applications is not realistic but the neural interconnections which code statistical interdependencies develop during experience (see, e.g., [57, 59]).

6.3. Preservation of redundancies have to be taken into consideration for the understanding of early visual processes

There are two main approaches to designing feature spaces for visual systems. The first is to *learn* features from natural images (see, e.g., [4, 6, 19, 28, 56, 67, 78]). The second approach is to define feature spaces explicitly as done in most technical applications (see, e.g., [53, 63, 74]). Related to this issue is the problem of innate versus learned structures in the human visual system[7]. Developmental psychology and neuro-physiological research give indications for the impressive adaptivity of the visual system and its capability to extract significant information from experience (see, e.g., [8, 68]). However, they also indicate a large amount of genetic pre-structuring [23, 38]. The connections of brain areas, the receptive field size of neurons in different areas and the topographic organization of these areas are largely predetermined and established at or soon after birth (see, e.g., [60]). Even the features extracted in some areas (orientation, movement and colour [23, 72]), i.e. the coarse sensitivity of neurons and their organization in feature maps, is basically initiated before the first post-natal visual experience[8]. We favour neither pure feature learning nor explicit feature design (actually we think the truth might be somewhere in between) but we want to discuss the consequences of our results in the context of the design of feature spaces in general.

An explicit definition of a feature space (as done in this paper) faces the problem that *a priori* assumptions about the structure of input signals have to be made. There is always a certain amount of arbitrariness involved in such settings and a misleading choice may lead to irrecoverable loss of performance. However, explicit definitions of feature spaces can be justified not only by neuro-physiological facts (see above) but also by conceptional needs since learning is inherently faced with the bias/variance dilemma (see, e.g., [22]): if the starting configuration of the system has many degrees of freedom, it can learn from and specialize to a wide variety of domains, but it will in general have to pay for this advantage by having many internal degrees of freedom—the 'variance' problem. On the other hand, if the initial system has few degrees of freedom it may be able to learn efficiently but there is a great danger that the structural domain spanned by those degrees of freedom does not cover the given domain of application at all—the 'bias' problem. Built-in *a priori* knowledge in terms of design decisions for feature spaces may be useful (and possibly necessary) for the creation of efficient visual systems since it allows the system to concentrate on significant aspects of the data and it is likely that such kinds of feature processing have been designed to a certain

[7] However, innate structures are learned as well, although by phylogeny and not by ontogeny.

[8] A very compact summary of the results of neurophysiology and developmental psychology concerning the role of experience and *a priori* knowledge in the human visual system is given in [47].

extent by evolutionary learning (see, e.g., [22, 47]). In our multi-modal feature space (which is motivated by analogies to the primate's visual system as well as by computer vision) we can show that significant redundancies corresponding to the gestalt principles of collinearity and parallelism can be found and we have argued how these redundancies can be used in an artificial system (see section 6.2) to overcome the ambiguity of local information.

An alternative approach to an explicit definition of feature spaces is the learning of features (overviews are given in [67, 78]). This has so far only been done for orientation selective cells with structure similar to Gabor wavelets [6, 56], but the future perspective is to extend this approach to intermediate and higher stages of visual processing [67] and to learn other more complex features. Many contributions indicate that a learning of early visual processing units can be guided and explained by statistical principles such as *redundancy reduction* [4], *independence* (see, e.g., [6]) or *sparse coding* (see, e.g.,[56]). Another useful criterion is *invariance* or *slow variation* [75] in fast varying input signals[9]. Our results suggest that the essential need to deal with uncertain data (see section 6.1) makes it likely that *redundancy preservation* might also be an important statistical principle in cortical processes to overcome the ambiguity of local feature processing (see also the discussion in [3, 66]). To put it in even stronger terms, redundancies in early vision are necessary for the applicability of gestalt principles and the applicability of modality fusion since the elimination of redundancies would disable any predictions between visual entities or across modalities at all (see also [3, 37, 57]).

An open question remains which *a priori* settings are sensible and what role learning has in the development of early and intermediate visual processing as opposed to or supplementing 'pre-wiring' mechanisms. Whatever answer one may prefer, we think the preservation of redundancies plays an important role for the learning of visual features as well as for the explicit design of feature spaces.

Acknowledgments

We especially thank Michael Felsberg for fruitful discussions and his support regarding the statistical investigations concerning the distribution of phases in natural images. Many thanks to Peter Kovesi for helping to generate figure 6 and to Peter Hancock for fruitful discussions and proofreading. We also would like to thank Matthias Henning, Hans Peter Mallot, Bill Phillips, Stefan Posch, Gerald Sommer, Laurenz Wiskott and Christoph Zetzsche for fruitful discussions.

Appendix A. Metrics for the different visual modalities

In this section we define metrics for the different modalities orientation, contrast transition, colour and optic flow.

Orientation

The orientation o takes values in the interval $o \in [0, \pi)$ (see figure 8(b), (i)). A metric $d^o(o, o')$ in the orientation subspace can be defined by

$$d^o(o, o') = \min\{|o - o'|, |(o + \pi) - o'|, |o - (o' + \pi)|\}$$

[9] All these studies demonstrate that the human visual system seems to establish a description of visual data in V1 which meets rather general requirements. This seems to be different in other species: e.g. frogs possess prey detectors in their retinas [51] and even among mammals different kinds of specialization exist in their retinal neurons. For example, in the rabbit retina motion detectors are found while in the primate's visual system such detectors first occur in V1 [52].

with $d^o(o, o') \in [0, \frac{\pi}{2})$. Since orientations close to 0 must have a small distance to orientations close to π the minimum over three cases has to be computed.

Contrast transition

The metric for contrast transition must also take the orientation into account, since a rotation of π corresponds to a switch of the sign of the phase. This leads to a slightly subtle definition of the metric. The phase takes values in the interval $p \in [-\pi, \pi)$ (see figure 8(b), (ii)). A metric $d^p(p, p')$ can be defined by

$$
\begin{aligned}
d^p(p, p') &= \min(d_1(p, p'), d_2(p, p')) \\
d_1(p, p') &= \tfrac{1}{2}(a\tan2(\sin(o - o'), \cos(o - o'))^2 + \tfrac{1}{2}(a\tan2(\sin(p - p'), \cos(p - p')))^2 \\
d_2(p, p') &= \tfrac{1}{2}(a\tan2(\sin(o - o' + \pi), \cos(o - o' + \pi))))^2 \\
&\quad + \tfrac{1}{2}(a\tan2(\sin(p + p'), \cos(p + p')))^2.
\end{aligned}
$$

The metric $d^p(p, p')$ takes values in $[0, \pi)$. An extended interpretation of this definition and the structure and meaning of the phase-orientation space will be given in [49].? See endnote 4 See endnote 5

Colour

The colour vector $\vec{c} = (c_r, c_g, c_b)$ (consisting of the averaged red, green and blue values in an image patch) takes values in $[0, 1] \times [0, 1] \times [0, 1]$. A metric in the colour space can be defined by

$$d^c(\vec{c}, \vec{c}\,) = \sqrt{((c_r - c'_r)^2 + (c_g - c'_g)^2 + (c_b - c'_b)^2)}$$

with $d^c(\vec{c}, \vec{c}\,) \in [0, \sqrt{3}]$. For a detailed discussion of metrics in different colour spaces see, e.g., [39, 43].

Optic flow

For the optic flow vector holds $(o_m, o_\alpha) \in [0, \infty) \times [-\pi, \pi)$, o_m representing the magnitude of the flow vector and o_α its angle (see figure 8(b), (iii)). A metric can be defined by

$$d^f((o_m, o_p), (o'_m, o'_p)) = \tfrac{1}{2}d(o_m, o'_m) + \tfrac{1}{2}d(o_\alpha, o'_\alpha)$$

with $d^f((o_m, o_p), (o'_m, o'_p)) \in [0, 2]$.

$$d(o_m, o'_m) = 1 - \frac{1}{1 + |o_m - o'_m|}$$

represents the distance in magnitude. The transformation $1 - 1/(1 + |o_m - o'_m|)$ transforms the differences in magnitude (which are in $[0, \infty)$) to the interval $[0, 1)$.

$$d(o_\alpha, o'_\alpha) = \frac{1}{\pi}\min(|o_\alpha - o'_\alpha|, |(o_\alpha + 2\pi) - o'_\alpha|, |o_\alpha - (o'_\alpha + 2\pi)|)$$

represents the distance in the angle coordinates. Since there is a periodicity at $-\pi$ and π (see figure 8(b), (iii)) the minimum of three cases has to be computed.

Appendix B. Normalization of the coordinate system according to e^2

Let $\mathcal{X}$ be our feature space $((x_1, x_2), o, p, ((c_r^l, c_g^l, c_b^l), (c_r^r, c_g^r, c_b^r)), (o_m, o_\alpha))$. We are after a transformation $T^{e^2} : \mathcal{X} \to \mathcal{X}$ which modifies the coordinate system such that the event $T(e^2)$ is in the origin with 0 orientation. This transformation is non-trivial since it does not only affect position and orientation but also phase, optic flow and colour (see figure 8(a)).

Given an entity $e^2 = ((x_1', x_2'), o', p', ((\vec{c}_l\,', \vec{c}_r\,'), (o_m', o_\alpha')))$ the transformation T^{e^2} is defined by

$$T^{e^2}((x_1, x_2), o, p, ((\vec{c}_l, \vec{c}_r), (o_m, o_\alpha))) = ((x_1 - x_1', x_2 - x_2'), T(o), T(p), T(\vec{c}), (o_m, T(o_p)))$$

with

$$T(o) := \begin{cases} o - o' & \text{if } o - o' \geqslant 0 \\ o - o' + \pi & \text{else} \end{cases}$$

$$T(p) := \begin{cases} p & \text{if } o - o' \geqslant 0 \\ -p' & \text{else.} \end{cases}$$

This holds since a rotation of π corresponds to a change of sign of the phase:

$$T(\vec{c}) := \begin{cases} (\vec{c}^{\,l'}, \vec{c}^{\,r'}) & \text{if } o - o' \geqslant 0 \\ (\vec{c}^{\,r'}, \vec{c}^{\,l'}) & \text{else.} \end{cases}$$

A rotation of π corresponds to the switch of the two colour patches:

$$T(o_\alpha) := \begin{cases} o_\alpha - o' & \text{if } o_\alpha - o' \geqslant 0 \\ o_\alpha - o' + 2\pi & \text{else.} \end{cases}$$

The optic flow rotates with the orientation but the singularity at $-\pi, \pi$ has to be taken into account.

Appendix C. Approximation of the gestalt coefficient

In this section we describe the approximation of the gestalt coefficient for two events from our data. Since we have to work with discrete samples of events it is very unlikely that we will find an exact identity $T(e) = e^1$ in our data set. Therefore we have to approximate this identity.

The numerator $P(e^1|e^2)$ in (1) can be defined by

$$P(T^{e^2}(e) \in \text{Bin}(e^1) \wedge d^i(T^{e^2}(e), e^1) < t^i|e^2) \tag{2}$$

with $i : p, c, f$ representing the different modalities phase (p), colour (c) and flow (f). We say $e \in \text{Bin}(e^1)$ when e and e^1 are close in the pixel-orientation domain, i.e. with $e = (x_1, x_2, o, \ldots)$ and $e^1 = (x_1', x_2', o', \ldots)$ it is

$$e \in \text{Bin}(e^1) \Leftrightarrow |x_1 - x_1'| < 5 \wedge |x_1 - x_1'| < 5 \wedge |o - o'| < \frac{\pi}{16}$$

according to the bin sizes of 10×10 and $\Delta o = \frac{\pi}{8}$ (see section 5)[10]. The term (2) represents the likelihood that a transformed entity $T^{e^2}(e)$ is close to the specific event e^1. Note that (as shown in figure 9) we sample space and orientation explicitly.

[10] Note that there is a difference between the modalities pixel position/orientation and the visual modalities contrast transition, colour and optic flow. The gestalt coefficient is measured for all pixel/orientation combinations (with the lower resolution defined by the bins). However, for contrast transition, colour and optic flow the gestalt coefficient is only computed for the event 'similar' (in the specific modality) to e^2, i.e. there is no explicit sampling of the full feature space.

Let e^r be an entity from our data set which is randomly transformed (the transformation parameters (x, y, o) are equally distributed)[11]. Then the denominator $P(e^1)$ can be defined by

$$P(T^r(e) \in \mathrm{Bin}(e^1) \wedge d^i(T^r(e), e^1) < t^i | e^r). \tag{3}$$

The term (3) expresses the likelihood that a *randomly* (equally distributed) transformed entity $T^r(e)$ is close to e^1.

In our data set we can now perform the following approximation:

$$\begin{aligned} P(e^1|e^2) &= P(T^{e^2}(e) \in \mathrm{Bin}(e^1) \wedge d^i(T^{e^2}(e), e^1) < t^i | e^2) \\ &\approx \frac{\#\{e | T^{e^2}(e) \in \mathrm{Bin}(e^1) \wedge d^i(T^{e^2}(e), e^1) < t^i\}}{N} \\ P(e^1) &= P(T^r(e) \in \mathrm{Bin}(e^1) \wedge d^i(T^r(e), e^1) < t^i | e^r) \\ &\approx \frac{\#\{e | T^r(e) \in \mathrm{Bin}(e^1) \wedge d^i(T^r(e), e^1) < t^i\}}{N}. \end{aligned}$$

#{ } is the number of elements in the set { } while N represents the number of entities e^2 in the data set. Note that, in the case of orientation only, the term $d^i(T^r(e), e^1) < t^i$ is always true and can therefore be neglected. In the case of regarding two modalities i, j at a time we have to apply two comparisons, i.e. $d^i(T^{e^2}(e), e^1) < t^i \wedge d^j(T^{e^2}(e), e^1) < t^j$.

References

[1] Aloimonos J and Shulman D 1989 *Integration of Visual Modules—An Extension of the Marr Paradigm* (London: Academic)

[2] Ayache N 1990 *Stereovision Sensor Fusion* (Cambridge, MA: MIT Press)

[3] Barlow H 2001 Redundancy reduction revisited *Network: Comput. Neural Syst.* **12** 241–54

[4] Barlow H B 1961 Possible principles underlying the transformation of sensory messages *Sensory Communication* pp 217–34 See endnote 6

[5] Barth E, Caelli T and Zetsche C 1993 Image encoding, labeling, and reconstruction from differential geometry *Graph. Models Image Process.* **55** 428–46

[6] Bell A J and Sejnowski T 1996 Edges are the 'independent components' of natural scenes *Adv. Neural Information Process. Syst.* **9** 831–7

[7] Bertenthal B I, Campos J J and Haith M M 1980 Development of visual organisation: the perception of subjective contours *Child Dev.* **51** 1072–80

[8] Blakemore C and Cooper G F 1970 Development of the brain depends on the visual environment *Nature* **228** 477–8

[9] Boyer K L and Sarkar S 1999 Perceptual organization in computer vision: status, challenges, and potential *Percept. Organiz. Comput. Vis.* **76** 1–5 (special issue)

[10] Brunswik E and Kamiya J 1953 Ecological cue-validity of 'proximity' and of other Gestalt factors *Am. J. Psychol.* **LXVI** 20–32

[11] Chung R C K and Nevatia R 1995 Use of monucular groupings and occlusion analysis in a hierarchical stereo system *Comput. Vis. Image Underst.* **62** 245–68 See endnote 7

[12] Cozzi A and Wörgötter F 2001 Comvis: a communication framework for computer vision *Int. J. Comput. Vis.* **41** 183–94

[13] Desolneux A, Moisan L and Morel J M 2001 Edge detection by the Helmholtz principle *J. Math. Imaging Vis.* **14** 271–84 See endnote 8

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

[15] Faugeras O and Robert L 1996 What can two images tell us about the third one? *Int. J. Comput. Vis.* **18**

[16] Faugeras O D 1993 *Three-Dimensional Computer Vision* (Cambridge, MA: MIT Press)

[17] Felsberg M and Sommer G 2000 A new extension of linear signal processing for estimating local properties and detecting features *Proc. DAGM 2000* pp 195–202 See endnote 9

[11] Since we are interested in the grouping of entities that occur within the very same image we make use of transformed entities extracted from an image instead of randomly generated entities that might not be close to our feature spaces (i.e. carry colour values that do not occur at all in the image).

[18] Felsberg M and Sommer G 2001 The monogenic signal *IEEE Trans. Signal Process.* **41**
[19] Field D 1987 Relations between the statistics of natural images and the response properties of cortical cells *J. Opt. Soc. Am.* **4** 2379–94
[20] Gazzaniga M S 1995 *The Cognitive Neuroscience* (Cambridge, MA: MIT Press)
[21] Geisler W S, Perry J S, Super B J and Gallogly D P 2001 Edge co-occurrence in natural images predicts contour grouping performance *Vis. Res.* **41** 711–24
[22] Geman S, Bienenstock E and Doursat R 1995 Neural networks and the bias/variance dilemma *Neural Comput.* **4** 1–58
[23] Gödecke I and Bonhoeffer T 1996 Development of identical orientation maps for two eyes without common visual experience *Nature* **379** 251–5
[24] Granlund G H and Knutsson H 1995 *Signal Processing for Computer Vision* (Dordrecht: Kluwer)
[25] Guy G and Medioni G 1996 Inferring global perceptual contours from local features *Int. J. Comput. Vis.* **20** 113–33
[26] Hahn M and Krüger N 2000 Junction detection and semantic interpretation using Hough lines *EIS' 2000*
[27] Hamming R W 1980 *Coding and Information Theory* (Englewood Cliff, NJ: Prentice-Hall)
[28] Hancock P J B, Baddeley R J and Smith L S 1992 The principal components of natural images *Network: Comput. Neural Syst.* **3** 61–72
[29] Heitger F, von der Heydt R, Peterhans E, Rosenthaler L and Kübler O 1998 Simulation of neural contour mechanisms: representing anomalous contours *Image Vis. Comput.* **16** 407–21
[30] Hoffman D D (ed) 1980 *Visual Intelligence: How We Create What We See* (Location: W W Norton and Company)
[31] Horn B K P and Weldon E J 1988 Direct methods for recovering motion *Int. J. Comput. Vis.* **2** 51–76
[32] Hubel D H and Wiesel T N 1979 Brain mechanisms of vision *Sci. Am.* **241** 130–44
[33] Ikeuchi K and Horn B K P 1981 Numerical shape from shading and occluding boundaries *Artif. Intell.* **17** 141–84
[34] Jähne B 1997 *Digital Image Processing—Concepts, Algorithms, and Scientific Applications* (Berlin: Springer)
[35] Jones J P and Palmer L A 1987 An evaluation of the two dimensional Gabor filter model of simple receptive fields in striate cortex *J. Neurophysiol.* **58** 1223–58
[36] Kalocsai P 1998 Contour completion algorithm quantitatively and qualitatively improves the performance of a recognition model (evidence against geon theory) *Proc. IEEE Computer Society Workshop on Perceptual Organization in Computer Vision*
[37] Kay J, Floreano D and Phillps W A 1998 Contextually guided unsupervised learning using local multivariate binary processors *Neural Net.* **11** 117–40
[38] Kellman P J and Arterberry M E (ed) 1998 *The Cradle of Knowledge* (Cambridge, MA: MIT Press)
[39] Klette R, Schlüns K and Koschan A 1998 *Computer Vision—Three-Dimensional Data from Images* (Berlin: Springer)
[40] Knudsen E I, du Lac S and Esterly S D 1987 Computational maps in the brain *Ann. Rev. Neurosci.* **10** 41–65
[41] Koch R 1994 Model-based 3D scene analysis from stereoscopic image sequences *ISPRS J. Photogr. Remote Sensing* **49** 23–30
[42] Kofka K (ed) 1935 *Principles of Gestalt Psychology* (New York: Harcourt and Brace)
[43] Koschan A 1994 How to utilize colour information in dense stereo matching and in edge based stereo matching? *Proc. ICARCV* pp 419–423
[44] Kovesi P 1999 Image features from phase congruency *Videre: J. Comput. Vis. Res.* **1** 1–26
[45] Krüger N 1998 Collinearity and parallelism are statistically significant second order relations of complex cell responses *Neural Process. Lett.* **8** 117–29
[46] Krüger N 1998 Collinearity and parallelism are statistically significant second order relations of complex cell responses *Proc. I&ANN 98*
[47] Krüger N 2001 Learning object representations using *a priori* constraints within orassyll *Neural Comput.* **13** 389–410
[48] Krüger N, Ackermann M, and Sommer G 2002 Accumulation of object representations utilizing interaction of robot action and perception *Knowl. Based Syst.* **13** 111–18
[49] Krüger N, Felsberg M, Gebken C and Pörksen M 2002 An explicit and compact coding of geometric and structural information applied to stereo processing *Proc. Workshop 'Vision, Modeling and VISUALIZATION' 2002*
[50] Ledgeway T and Hess R F 2002 Rules for combining the outputs of local motion detectors to define simple contours *Vis. Res.* **42** 653–9
[51] Lettvin J Y, Maturana H R, McCulloch W S and Pitts W H 1959 What the frog's eye tells the frog's brain *Proc. Inst. Radio Eng.* **47** 1950–61
[52] Mallot H A 2001 personal discussion

See endnote 10
See endnote 11
See endnote 12
See endnote 13
See endnote 14

[53] Mel B 1996 Seemore: a view-based approach to 3D object recognition using multiple visual cues *Adv. Neural Information Process. Syst.* **8** 865–71
[54] Mundy J L, Liu A, Pillow Nic, Zisserman A, Abdallah S, Utcke S, Nayar S and Rothwell C 1996 An experimental comparison of appearance and geometric model based recognition *Object Representation in Computer Vision* pp 247–69
[55] Nagel H-H 1987 On the estimation of optic flow: relations between different approaches and some new results *Artif. Intell.* **33** 299–324
[56] Olshausen B A and Field D 1996 Emergence of simple-cell receptive field properties by learning a sparse code for natural images *Nature* **381** 607–9
[57] Phillips W A and Singer W 1997 In search of common foundations for cortical processing *Behav. Brain Sci.* **20** 657–82
[58] Posch S 1997 *Perzeptives Gruppieren und Bildanalyse* Habilitationsschrift, Universität Bielefeld, Deutsche Universitäts Verlag
[59] Prodöhl C, Würtz R and von der Malsburg C 2002 Learning the gestalt rule collinearity from object motion *Neural Comput.* submitted
[60] Rakic P 1995 Corticogenisis in human and nonhuman primates *The Cognitive Neuroscience* ed M S Gazzaniga (Cambridge, MA: MIT Press) pp 127–45
[61] Rohr K 1992 Recognizing corners by fitting parametric models *Int. J. Comput. Vis.* **9** 213–30
[62] Sarkar S and Boyer K L 1994 *Computing Perceptual Organization in Computer Vision* (Singapore: World Scientific)
[63] Schiele B and Crowley J L 1996 Probabilistic object recognition using multi-dimensional receptive field histograms *Adv. Neural Information Process. Syst.* **8** 865–71
[64] Schmid C and Zisserman A 1997 Automatic line matching across views *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 666–71
[65] Schwartz O and Simioncelli E 2001 Natural signal statistics and sensory gain control *Nat. Neurosci.* **4** 819–25
[66] Sigman M, Cecchi G A, Gilbert C D and Magnasco M O 2001 On a common circle: natural scenes and Gestalt rules *Proc. Natl Acad. Sci. USA* **98** 1935–49
[67] Simoncelli E P and Ohlshausen B A 2001 Natural image statistics and neural representations *Annu. Rev. Neurosci.* **24** 1193–216
[68] Sur M, Garraghty P E and Roe A W 1988 Experimentally induced visual projections into auditory thalamus and cortex *Science* **242** 1437–41
[69] von der Malsburg C 1981 The correlation theory of brain function *Internal report*
[70] Watt R J and Phillips W A 2000 The function of dynamic grouping in vision *Trends Cognitive Sci.* **4** 447–154
[71] Wertheimer M (ed) 1935 *Laws of Organisation in Perceptual Forms* (London: Harcourt Brace Jovanovich)
[72] Wiesel T N and Hubel D H 1974 Ordered arrangement of orientation columns in monkeys lacking visual experience *J. Comp. Neurol.* **158** 307–18
[73] Willshaw D J (ed) 2001 *Network: Comput. Neural Syst.* **12** (Special issue)
[74] Wiskott L, Fellous J M, Krüger N and von der Malsburg C 1997 Face recognition by elastic bunch graph matching *IEEE Trans. Pattern Anal. Mach. Intell.* **19** 775–80
[75] Wiskott L and Sejnowski T 2002 Slow feature analysis: unsupervised learning of invariances *Neural Comput.* **14** 715–70
[76] Wörgötter F, Cozzi A and Gerdes V 1999 A parallel noise robust algorithm to recover depth information from radial flow *Neural Comput.* **11** 381–416
[77] Wuescher D M and Boyer K L 1991 Robust contour decomposition using constant curvature criterion *IEEE Trans. Pattern Anal. Mach. Intell.* **13** 41–51
[78] Zetzsche C and Krieger G 2001 Nonlinear mechanisms and higher-order statistics in biological vision and electronic image processing: review and perspectives *J. Electron. Imaging* **10** 56–99

See endnote 15
See endnote 16
See endnote 17