Technical Report SFB360-TR-98/10

# Perceptual Grouping using Markov Random Fields and Cue Integration of Contour and Region Information

Stefan Posch*, Daniel Schlüter
Applied Computer Science
University of Bielefeld

**Abstract**

A common feature of computer vision systems is the use of levels of increasing abstraction to represent intermediate results and thus successively bridging the gap between raw image data and the final result. To elaborate on such a hierarchical representation we propose a contour-based grouping hierarchy based on principles of perceptual organization. Exploiting regularities in the image, we aim at enhancing efficiency and robustness of subsequent processing steps of an image analysis system by reducing ambiguities in the intermediate representation and by realizing image primitives at a higher level of abstraction. To this end, first grouping hypotheses are generated within the hierarchy using the concept of areas of perceptual attentiveness. Since the generation is based on local evidence only, the hypotheses have to be judged in a global context. We employ a Markov Random Field to model context dependencies and energy minimization yields a consistent interpretation of image data with groupings from the hierarchy. Since this grouping hierarchy is contour-based, it inherits from drawbacks of contour segmentation. Therefore, the second issue we address aims at the integration of cues from region segmentation into the contour-based grouping process and vice versa. This integration is done on the level of contour segments, collinearities, and curvilinearities with complete regions to support and constrain each other. Results are presented for real images from a construction domain.

## 1 Introduction

The goal of image analysis is to derive a (symbolic) description of image data, which is suitable for a given task and will often comprise objects detected in the scene along with their attributes and relations. A common feature of computer vision systems is that this transformation is accomplished within a hierarchy of abstraction levels, beginning

*email: posch@techfak.uni-bielefeld.de



Technical Report 98/10, SFB 360, Universitaet Bielefeld, 1998.

with the raw sensor data up to the desired symbolic description, where the control flow does in general not have to agree with the hierarchy. The main goal of this hierarchy is to explicitly represent the information inherent in the data with growing abstraction, making also use of general or domain specific knowledge to a varying degree. Thus, the gap between raw image data and high level representations like object models is bridged successively giving rise to a number of more simple modules interacting to achieve a perceptual ability.

To elaborate on such a hierarchical representation is the first focus of our work. Starting with a contour segmentation of a static image, we exploit further regularities in the image using rather general, domain-independent assumptions about the projected objects to organize contour segments making use of principles of perceptual organization (or grouping) as put forward by the Gestalt psychologists (see e.g. [44]). Initially, this aims at overcoming fragmentation of contours due to noise and occlusion, but also at representing image data at a higher level of abstraction using parallel and closed structures. Summarizing, the ultimate goal of the grouping process is to enhance efficiency and robustness of subsequent processing steps of an image analysis system, like object recognition, by reducing ambiguities in the intermediate representation and by realizing image primitives at a higher level of abstraction.

The second issue we are concerned with is cue integration of contour- and region-based segmentation. Both segmentation techniques are dual in the sense, that each region is bounded by a closed contour and vice versa. However, due to the different views on segmentation, their results exhibit different advantages and drawbacks. Similar to biological vision systems (see e.g. [8, 40]) we start with separate pathways for contour and region segmentation processing the image data independently. Both segmentations are combined at the level of contour segments and complete regions to yield additional contour and region interpretations for the image data using information of the other channel. Thus, both segmentation results are mutually enhanced and extended to improve further processing.

This work is embedded into a joint effort of researchers in the field of pattern recognition, artificial intelligence, and linguistics to study advanced human-machine interaction (see [35] for an overview). The domain was chosen to be the cooperative construction of a toy-airplane with parts from a wooden construction-kit for children and the examples presented in this work are drawn from this domain.

# 2 Related Work

The ideas of perceptual grouping for computer vision have their roots in the well known work of Gestalt psychologists back at the beginning of the century who described, among others, the ability of the human visual system to organize parts of the retinal stimulus to "Gestalten", into organized structures (see e.g. [44, 36]). The importance of perceptual grouping for computer vision has first been recognized by Witkin and Tenenbaum [45] and Lowe [24]. They introduce the principle of non-accidentalness stating that structure and regularity in the image is likely to occur due to an underlying cause in the three dimensional scene, since regularity is preserved when projecting a structured world.

Therefore, detecting regularities in the image indicates a common cause in the scene with a high probability.

In the last decade these ideas have drawn increasing attention in the computer vision community. Sarkar and Boyer propose in [38] a scheme to classify grouping algorithms according to the dimensionality and complexity of image primitives (or features) they are working with. The latter dimension is divided in four layers of abstraction: signal, primitive, structural, and assembly level, where the border between levels cannot be defined precisely. Perceptual organization within the signal level essentially refers to image segmentation for static images and motion detection for image sequences which usually is not subsumed under the term of grouping. However, including them into the classification scheme explicitly states the common principle of organizing entities according to some common properties.

Most of work in perceptual grouping has been done in the primitive and structural level for static 2D images. The focus of our work presented in this paper is in the latter one and we concentrate on this area in the following. Besides the abstraction level and the image dimension as reflected in the scheme of Sarkar and Boyer the representational and algorithmic techniques employed by a grouping method is a main source for discrimination (which deliberately has not been included in the classificatory structure by the authors).

The primitive level is mainly concerned with one dimensional processes. For contour-based approaches this essentially equals to extending (short) segments or to organize pixels or edgels into salient extended contours. Work from this class of grouping includes [9] where initial edgels are grouped using the principles of proximity and good continuation to form strands. These are classified as straight lines, inflections, cusps, corners, or conics and replaced by new tokens. With the most significant of these new tokens the process is recursively iterated on a coarser spatial scale. This system is an extension of [4] concerned with grouping into straight line segments. In both the work of [41, 39] and [12], first a measure of saliency for all pixels in the image is computed and subsequently contour segments are derived from these saliency maps. While in the former system, the saliency map is computed recursively with local optimization favoring smooth and elongated contours, the latter uses the concept of extension fields and employs a directional convolution at each image site. The approach in [42, 43] groups edge elements from initial edge detection into straight line segment using a Markov Random Field. The edgels are defined as the sites and their labels are composed as the triple of the index of an extended segment, the direction, and the state (e.g. termination or no edge). For initialization of the index, neighboring edge elements with common privileged direction are assumed as initial straight line segments and assigned a common label. The various clique potentials model the data dependency, a tendency to extend a straight line taking its current direction into account, and the configuration of number and states of edge elements within a 8-neighborhood.

In contrast to the primitive level, methods within the structural level also take two dimensional grouping principles into account, often with larger spatial extent. The work of Lowe [24] falls into this category. To group straight line segments he employs parallelism besides collinearity and proximity. Grouping is based on a significance reflecting the a priori probability of its occurrence, where a circular neighborhood of the segments is used for this computation. Similar approaches are proposed in [30] using semantic networks, in [48] with symbolic rules, and in [17] using Fuzzy sets.

There are several approaches representing contour segments and their relations in a graph and accomplish organization of these segments with various algorithms on this graph: In the generic grouping method proposed in [1], grouping cues which judge the consistency of a given set of image primitives to belong to a common group are represented as binary relations by edges in this graph connecting image primitives as nodes. To this end, a cue enhancement procedure is proposed to incorporate higher order cues when deciding for these binary relations. Within this graph the groupings are found as cliques and are computed with a heuristic algorithm to approximate a maximum likelihood criterion. In [14], constraint triangulation is used to constructed a graph from straight line segments to describe neighborhood relations between segments. A cost function is defined on the nodes of this graph to reflect the quality of local convexity. From this graph, open and closed convex groups of straight line segments are easily extracted. To detect generic closed contours, an initial segmentation for straight line segments is used again in [13]. For each pair of two segments a score is defined taking geometric and photometric attributes into account, where segments are considered to be directed. After selecting locally significant interpretations, these are represented in a graph and non-intersecting cycles describe closed contours. In addition to straight line segments, the system described in [16] uses projections of orthogonal edges to delineate buildings in real images. These low level image primitives are stored in a graph as nodes and are connected with an edge given that smoothness and symmetry conditions hold. Then closed contours are searched for in this graph as cycles, where evidence for missing corners to close a path is gathered from image data running a corner detection algorithm in a model-driven way. To disambiguate between conflicting hypotheses of closed contours, all cycles are represented in a second graph as nodes and conflicts as edges. The final interpretation is derived scoring closed contours with their mean contrast and searching for non-conflicting nodes with maximal score in the second graph. The system in [15] aims at detecting closed contours of different shape type starting from contours segments of smooth curvature. All segments are processed sequentially, ordered by their saliency defined using length and degree of smoothing. Each segment under consideration is included into the image interpretation, if it significantly contributes information to the geometrical configuration as described by the already selected segments. To this end, tests on proximity, collinearity, curvilinearity, and parallelism are employed and junctions between contour segments are introduced. All accepted segments are represented in a rigid graph and minimal cycles detected as image regions. Finally, these are classified into a shape alphabet on the basis of their curvature signature.

In approaches using Markov Random Fields, graphs are used implicitly to define the underlying neighborhood system. In the system [19] to delineate buildings in areal images, straight line segments constitute the sites of a random field which are to be labeled with a significance. The neighborhood for each segment is defined using spatial proximity, orientation difference, and the currently assigned significance, thus giving rise to a dynamic evolving neighborhood system. The clique potentials model length and geometric configuration, as well as an entropy term to drive labels to either 0 or 1. The significance is the basis for the final selection or rejection of a segment. Boundaries from selected line segments are finally approximated using active contours. MRFs are also used in [29] for grouping straight line segments into disjunct sets. Again the nodes of a graph are defined as the line segments with an associated label indexing the group it is an element of. The grouping principles collinearity, parallelism, and convergence are represented as multi-dimensional edges which define independent neighborhood systems on the graph. Potentials for cliques with two elements are used to score the compliance to these grouping principles and are combined to a global energy function. This is minimized using simulated annealing yielding a partition of line segments as a result.

A special feature of the grouping technique proposed in [26] is the uniform representation of contour- and region-based image primitives with bounding envelops, which are identical to alpha shapes as defined in [10]. For each set of grouping primitives the principles of good continuation, proximity, similarity, and closure are modeled as six different energy terms with their linear combination as the resulting energy. Organizing image primitives of an initial segmentation into disjunct sets of primitives is accomplished as energy minimization over all possible partitions with simulated annealing.

While all these approaches make use of grouping principles like symmetry or closure, the entities represented are one dimensional in character, with the nodes of the second graph in [16] as the only exception. In contrast, the last two grouping systems to be outlined explicitly model two dimensional structures as groupings in their framework. In [27, 28] contour segments are organized into collated features, and these in turn into parallels, u-contours, and rectangles. A constraint satisfaction network accomplishes the global judgment using supporting and competing relations between these grouping primitives. In a second similar system, more general shapes are considered defining a measure for symmetry between polygons and symmetric contours forming ribbons. In [38] the grouping process is organized into different levels of a hierarchy. At each level of this hierarchy voting methods and graph algorithms implement a preattentive module to generate grouping hypotheses taking grouping principles like proximity or similarity into account. For the subsequent attentive phase, Bayesian networks are extended to Perceptual Inference Networks to efficiently code knowledge about structural dependencies between groupings and use this to infer missing data. Conflicts between resulting hypotheses are resolved in a final step using conflict graphs and special purpose modules.

Our approach to grouping uses a similar framework as the last two system which will become apparent in the next section. In addition to a contour-based view, we also incorporate information from region-based segmentation an vice versa. Also the human visual system uses different types of visual cues for perception. Regarding computer vision, Sanocki, Bowyer, Heath and Sarkar [37] discuss the question, "Are edges sufficient for object recognition?". Their conclusion is, that contour information alone is not sufficient and they propose to use other sources of information like regions, surfaces or intermediate-level structures beside edge information to constrain each other. To combine region- and edge-based techniques several algorithms have been proposed. Pavlidis and Liow [32] apply criteria regarding contrast, gradient variation, and artifacts along region boundaries to verify and eliminate boundaries obtained from an over-segmenting split-and-merge method. In a similar way, [47] uses the results of edge detection to verify and correct region boundaries generated by region-growing. The approach of [46] integrates gradient information into a region growing process. In contrast, [31] proposes a method to create a fully connected line-drawing from an edge segmentation using a stochastical region growing. Other approaches using similar combinations of region and edge segmentation can be found in [2] and [18]. While these approaches employ different region and edge detection techniques, all of them rely on pixel-based combinations of segmentation results. We propose to combine both cues at the level of contour-based groupings and complete regions to make use of more abstract image information.

In the following section an overview of our approach is given and the outline of the paper is given in more detail at the end of this section.

# 3 Overview

In the following we propose a contour-based approach to perceptual grouping. The main goal is to make explicitly available the implicit structure of the image with mechanisms of perceptual organization. A contour based approach as described here is not able to make use of every Gestalt principle. Thus we concentrate on proximity, good continuation, symmetry, and closure relations. Good continuation is referred to as collinearity for line segments and curvilinearity for elliptical arcs and parallelism is used as a special case of symmetry. Closure assumptions are formed from proximities or symmetry relations. To overcome inherent drawbacks of a contour-based technique, we additionally incorporate region information into this grouping system, and inversely enhance region segmentation with the results of perceptual organization.

## 3.1 Initial Segmentation

For initial segmentation, we first extract edge elements with the standard Sobel operator and with subsequently thinning using non-maximum suppression and thresholding with hysteresis (see e.g. [5]). After edge detection the edge elements are approximated with straight lines and elliptic arcs with a technique due to Leonardis (see [22, 21]). These contour models yield an adequate description of the 2D projections of surface boundaries of objects in our domain. However, contours representing surface boundaries are often fragmented due to possible occlusions of different objects or illumination effects and one aim of the grouping process is to overcome these fragmentations. The resulting contour segments are the basis for the definition of the proposed grouping hierarchy as shown in Figure 1 and the generation of hypotheses within this hierarchy. If a richer set of contour models is necessary, the grouping hierarchy has to be adapted accordingly.

## 3.2 Grouping Hierarchy

To model grouping phenomena of varying spatial complexity we introduce a hierarchy of grouping hypotheses with increasing complexity. Different levels of the hierarchy are distinguished by the type of image primitives or groupings and by the Gestalt principles

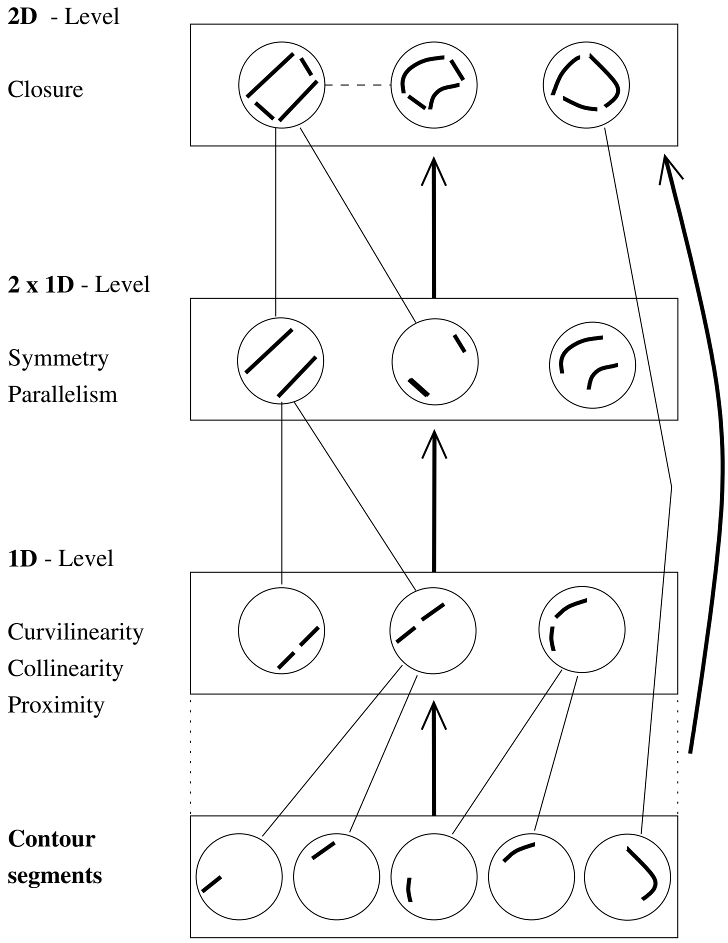



Figure 1: Hierarchy of grouping hypotheses: Each group is depicted as a circle, some examples of neighborhood relations are given as thin solid lines (supporting) and thin dashed lines (competing). Thick arrows show the data flow of the generation of hypotheses according to the hierarchy. The dotted lines connecting the 1D level and initial segmentation indicate, that contour segments are considered as primitive groups.

employed. Figure 1 shows that the *1D-Level* uses the Gestalt principles of good continuation and proximity to group contour segments. This first level of the hierarchy is called 1D-Level, since it deals with one dimensional phenomena. It introduces three different types of grupings. The first is the *collinearity* which is built from two or more straight line segments. This yields a new interpretation of the underlying edge elements of these line segments, again as a straight line. Analogously, two or more elliptical arcs may be combined to from the second grouping type within the 1D-Level, a *curvilinearity*, which is approximated once more by an elliptical arc. Alternatively, curvilinearities could be approximated by more complex contour models or represented without an parametric approximation at all.

Some contours of the initial segmentation cannot grouped into collinearities or curvilinearities if the already represent a complete interpretation of the underlying image data with respect to the principle of good continuation. These contour segments are called *primitive groups* and are treated equivalently to collinearities or curvilinearities. Which of the initial contour segments has to be interpreted as a primitive group in this sense has to be decided after final judgment of all hypotheses. In Figure 1 this is expressed by connecting the basis of contour segments with the 1D-Level of the hierarchy. Furthermore, we introduce the term *linear group* for all contour segments, collinearities, and curvilinearities.

The third type of grouping defined within the 1D-Level employs the principle of proximity. We will group always two linear groups to form a *proximity (group)*. Proximities play a special role compared to all other types of groupings, since they are not treated as interpretation of image data in the sense of image primitives. This would result in a redundant interpretation, and proximities are therefore viewed as relations between linear groups subsequently.

The next level of the hierarchy exploits the principle of symmetry and parallelism to group two linear groups. The resulting *symmetric* or *parallel group* is interpreted as part of the same surface boundary or surface texture. Since this extents a one dimensional view, but does not describe a closed image area, we name it the $2 \times 1D$ *-Level*.

Finally, the *2D-Level* treats closed figures using the principle of closure. Hypotheses of the lower level are used to construct a *closedness* grouping. We distinguish two types of closedness groupings. The first one is built from a closed sequence of linear groups which are pairwise neighboring according to proximity relations from the 1D-Level. The second one also incorporates parallel groups between the involved linear groups thus introducing additional evidence of an underlying cause.

Throughout this paper we use the following notation for the different types of groupings:

| | |
|---|---|
| $g_1 \| g_2 \| \dots \| g_n$ | collinear group |
| $g_1(g_2(\dots(g_n$ | curvilinear group |
| $g_1 \cdot g_2$ | proximity |
| $g_1 \\| g_2$ | parallel group |
| $g_1\{g_2$ | symmetric group |
| $g_1 \circ g_2 \circ \dots \circ g_n$ | closedness |

For all types of groupings we define the *grouping elements GElem* as the parts of each group, e.g. $GElem(g_1|\dots|g_n) = \{g_1, \dots, g_n\}$.

The various groupings are not isolated within the hierarchy, but are related to each other as illustrated by Figure 1. The construction of the groups yields the first type of relation in a quite natural way: All but the primitive groups consist of a set of grouping elements resulting a part-of relations. The second type of relation is motivated by the observation, that different groups may yield contradicting interpretations of the underlying image data. It will be defined more rigorously in Section 6.

This hierarchy has similarities with both the methods of Mohan and Nevatia (see [27, 28]) and the one of Sarkar and Boyer (see [38]). Compared to the former, our proposed hierarchy contains more descriptional power due to more general proximities and closedness hypotheses and facilitates the representation of transitive part-of relations not explicitly rather implicitly by the judgment procedure. All grouping types of our hierarchy are represented in the 1D-Level of the hierarchy as defined by Sarkar and Boyer. Since they model dependencies between grouping hypotheses with Bayes nets, their relations are directed and only supporting relations are represented. Contradicting interpretations have to be resolved in a subsequent step using a conflict graph, while in our approach this is integrated within the judgment procedure.

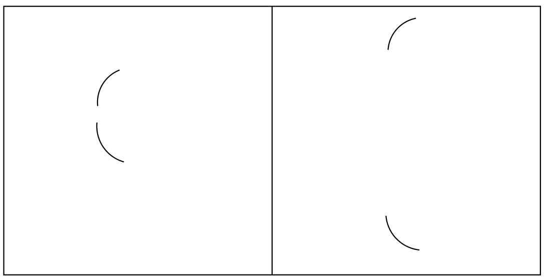
In the next section, an outline of the process to generate grouping hypotheses within the hierarchy is given. Cue integration of results from region segmentation with contour-based grouping hypotheses is described in the subsequent section. Our model to use Markov Random Fields to model the dependencies between hypotheses is presented in Section 6. Finally, results are given in Section 7 and a summary is given in the concluding section.

# 4 Generation of Grouping Hypotheses

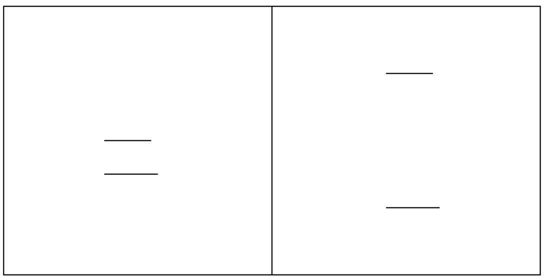
The first stage of the grouping process is to generate grouping hypotheses. Hypotheses are constructed bottom up with respect to the grouping hierarchy implementing the various Gestalt principles to organize the image data. This generation is based on local evidence only, where we make use of areas of perceptual attentiveness. We first introduced this concept and how the areas are derived from a trainging set. Subsequently, the process generating hypotheses is described in detail in the rest of this section. Judgment within a more global context to resolve ambiguities of this interpretation is deferred to the second stage (see Section 6).

## 4.1 Areas of Perceptual Attentiveness

Distance between contour segments is of importance not only for proximity groupings, but also for collinearity, curvilinearities, and parallel groups as illustrated in Figure 2. Grouping will only occur within a certain distance depending on the size of the involved

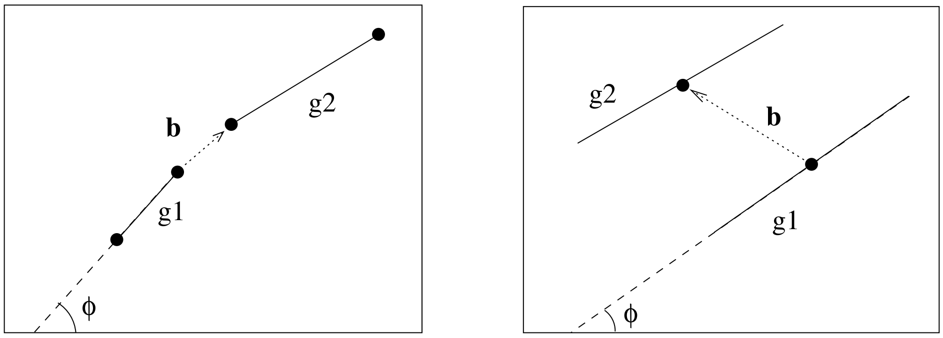


(a)



(b)

Figure 2: The strength of grouping effects according to the principle of good continuation and parallelism decreases with the distance between parts: In both subfigures the image primitives are of the same size and orientation, only the distance has been varied.




(a) (b)

Figure 3: Normalization of relative reference point positions for (a) collinearities and (b) parallel groups. Reference points are depicted with solid black circles. Given the current group $g1$ and its manually grouped partner $g2$, the relative position of reference points is denoted by $\mathbf{b}$. To normalize with respect to orientation of $g1$ we use rotation angle $\phi$.

image primitives, among others[1]. This fact is modeled as a search area for each image primitive restricting the relative location of potential grouping partners. We call these search areas *Areas of Perceptual Attentiveness* thus taking an active view of image primitives. Information about shape and size of these areas is derived from a hand labeled training set of our domain for each grouping principle considered. For the process of organizing these areas are used to delimit the search area after proper scaling and normalization as described in Subsection 4.2.

After initial segmentation a sample set of 40 images has been grouped manually to yield all intended significant groupings for the 1D and $2 \times 1$D level of the hierarchy. Groups of the 2D level have not to be considered, as will become clear later. This labeling is analyzed for each type of grouping hypotheses regarding the distribution of the relative position of reference points between the grouped items. As reference points we use endpoints for linear groupings of the 1D level. For parallel and symmetric groups the center points are defined as reference point (see Figure 3 for an illustration). For each given contour $g1$ we compute the relative position $\mathbf{b}$ of its reference point to all manually matched linear groups $g2$. This is done for both reference points of each linear and parallel group. Since search areas are not generally circular, we analyze relative reference point position taking the orientation of the group $g1$ into account. In addition, normalization of the distance with respect to the length of linear groups is necessary as has been illustrated in Figure 2. Thus, evaluation of the training set results in a distribution of normalized reference point positions like those shown in Figure 4.

From this distribution we derive connected areas of perceptual attentiveness by first blurring the distribution with Gaussian kernels, where the width increases with relative frequency at a given position. Subsequently, we apply thresholding and morphological closing, and describe the resulting areas by their convex hull shown in Figure 5. For

[1]For example, also further image primitives in between will effect the strength of grouping principles which we do not consider at this point.

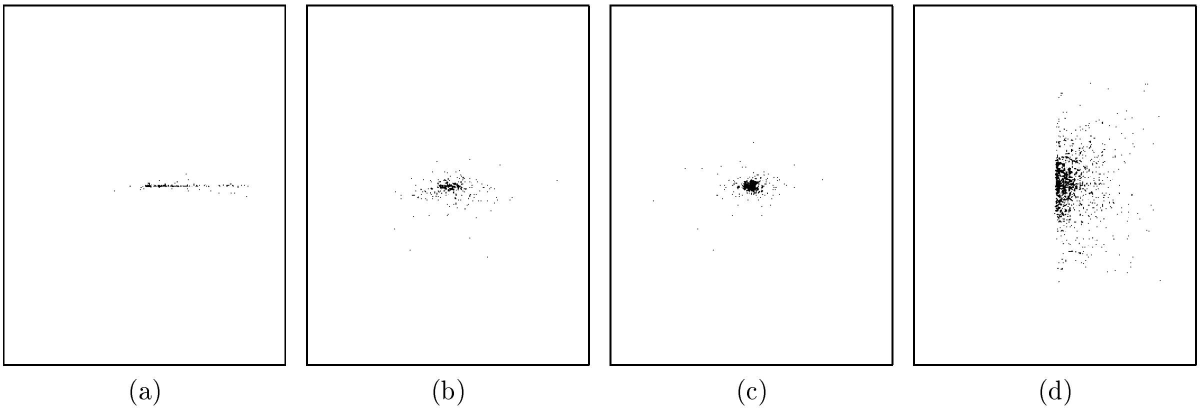


Figure 4: Frequence of relative reference point positions for (a) collinearity, (b) curvilinearity, (c) proximity, and (d) parallelism. The reference point of the current group $g1$ (see figure 3) is located in the center of the frame.

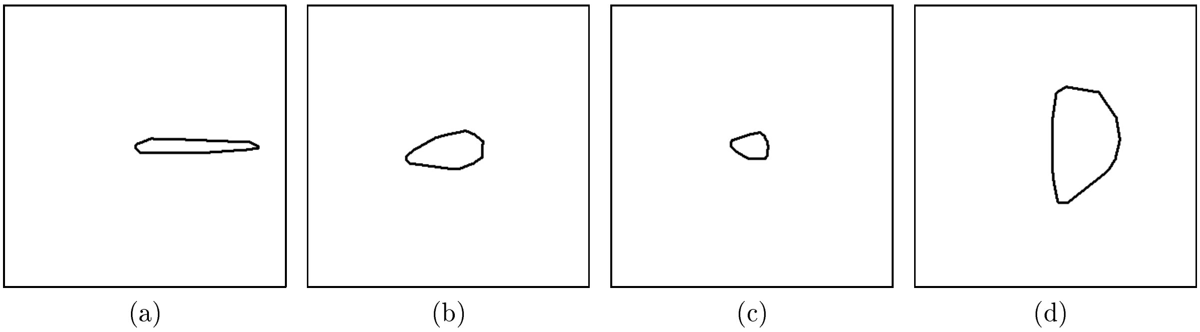


Figure 5: Areas of perceptual attentiveness as derived from the training set for (a) collinearity, (b) curvilinearity, (c) proximity, and (d) parallelism.

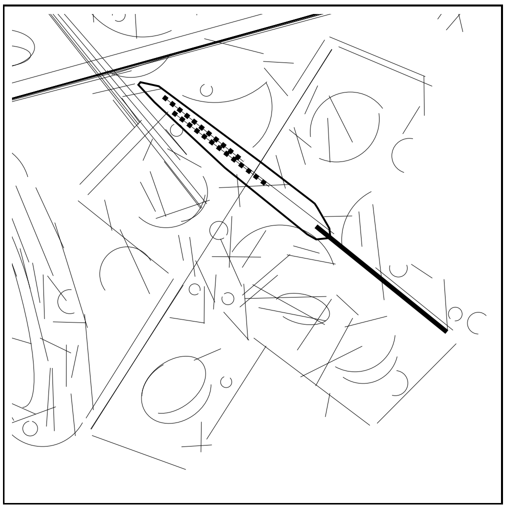
more details see [25, 33]. These areas of perceptual attentiveness are now employed for generating grouping hypotheses within a local image context.

## 4.2 1D and $2 \times 1$D level

In this section we formulate Gestalt principles for the different types of groupings in a precise way to facilitate the generation of grouping hypotheses. For two straight line segments to form a **collinearity**, the following two conditions have to be met:

1. The line segments have to lie approximately on a straight line and
2. the gap in between has to be small compared to the length of segments.

To precisely formulate the second requirement, we employ the area of perceptual attentiveness derived for groupings of type collinearity. Before applying, it is first scaled and rotated according to the size and orientation of each individual segment exactly as was done for evaluating the training set (see Figure 6 for an example). In addition the orientation difference has to be below a threshold $\Delta_{\mathrm{col}}$ to meet the first condition. This



(a)

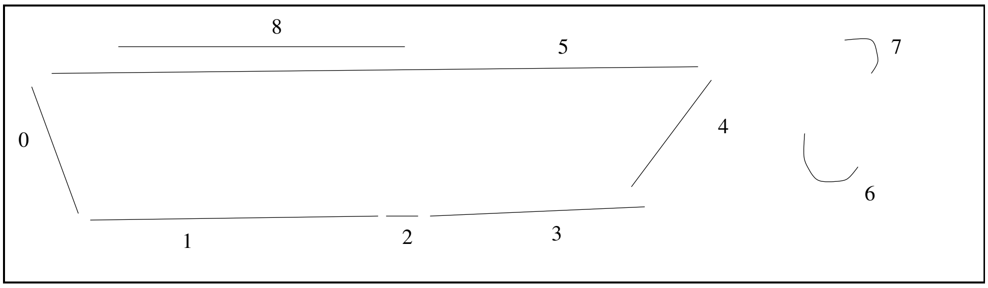
Figure 6: Example for the application of the area of perceptual attentiveness for a grouping of type collinearity. The mask has been rotated and scaled according to the orientation and size of the current line segment, which is displayed as a thick line. Segments used to generate collinearities are shown as dotted lines.

threshold again is derived from the training set and given in Table 1. If both requirements are satisfied, we generate a hypothesis for a collinearity, and the underlying edge elements of the grouped contour segments are again approximated with a straight line segment. After grouping, this new linear group recursively takes part in the generation process. Therefore, collinear groups with three or more underlying contour segments of the initial segmentation can be hypothesized (see $g1|g2|g3$ in Figure 7(a) as an example). In this case, due to the size of the search areas, the collinearity $g1|g3$ is first hypothesized in addition. This is however a redundant group since both interpret the same image data in the same way. Therefore we prefer the first linear group as a hypothesis, because it contains more contour segments of the initial segmentation, and to this end formally define redundancy:

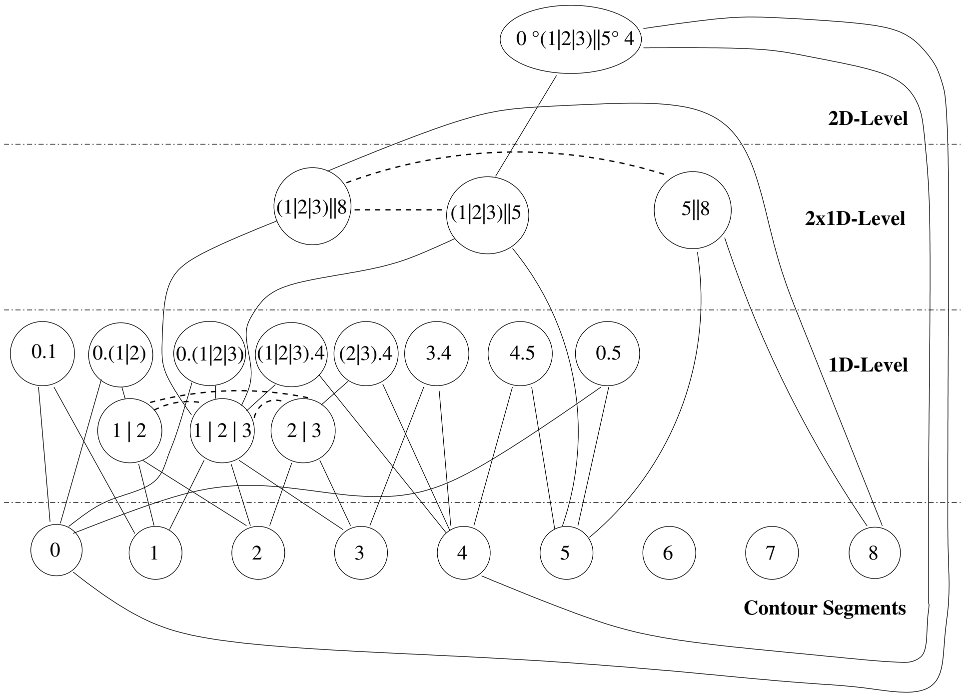
> Let $g = g_1|g_2|\ldots|g_n$ and $\tilde{g} = \tilde{g}_1|\tilde{g}_2|\ldots\tilde{g}_m$ be two collinear groups. Then $g$ is defined to be *redundant* with respect to $\tilde{g}$, iff
>
> $$((g_1 = \tilde{g}_1 \wedge g_n = \tilde{g}_m) \vee (g_1 = \tilde{g}_m \wedge g_n = \tilde{g}_1)) \wedge \\ GElem(g) \subset GElem(\tilde{g}) \tag{1}$$

| | |
|---|---|
| $\Delta_{\text{col}}$ | 2.9° |
| $\Delta_{\text{curvi}}$ | 17.2° |
| $\Delta_{\text{par}}$ | 2.8° |
| $\Delta_{\text{par\_length}}$ | 0.3 |

Table 1: These thresholds for the generation of grouping hypotheses have been derived from the training set of 40 images.




(a)




(b)

Figure 7: Example for the generation of grouping hypotheses and the construction of the grouping graph: (a) Synthetic segmentation and (b) grouping graph.

Each collinear group being redundant according to this definition is discarded from the set of hypotheses.

Hypothesizing of **curvilinearities** follows the same scheme. We require that

1. two elliptical arcs lie approximately on an ellipse and
2. the gap in between has to be small compared to the length of the segments.

For the first condition the orientation difference of the tangents at the endpoints is computed. It has to be below a threshold $\Delta_{\text{curvi}}$ degrees, which again is derived from the training set (see Table 1), and the signs of the curvature have to agree. The area of perceptual attentiveness is again used to define the second requirement. As for collinear groups, a curvilinear group is approximated by an elliptical arc and grouping is iterated recursively. Also, redundancy is defined and handled analogously.

For two linear groupings to form a **proximity** only one condition has to be met. (Recall, that a linear group has been defined to be a contour segment, a collinearity, or a curvilinearity.)

1. The gap in between has to be relatively small.

Therefore, we apply the area of perceptual attentiveness to all linear groups, thus allowing proximity groupings e.g. between a collinearities and an elliptical arc. If two linear groups have already been grouped according to the principle of collinearity or curvilinearity, they are not considered for a proximity grouping, to avoid redundant groupings.

Two line segments of collinearities are hypothesized as a **parallel group**, if

1. they are approximately parallel,
2. overlap, and
3. are spatially relatively close.

The first condition is defined as a threshold on the orientation difference $\Delta_{\text{par}}$ degrees. The overlap for two collinear groups $g1$ and $g2$ with length $l(g1)$ resp. $l(g2)$ is approximate by

$$\frac{min(l(g1), l(g2))}{max(l(g1), l(g2))} \tag{2}$$

and has to be below the threshold $\Delta_{\text{par\_length}}$. Both threshold are again determined from the training set (see Table 1). The requirement for spatial proximity is again concretized by applying scaled and normalized areas of perceptual attentiveness.

Again, redundant hypotheses can be constructed comparing all pairs of collinear groups and line segments. Consider again the synthetic example in Figure 7(a). Both groups $g_1|g_2$ and $g_1|g_2|g3$ initially form a parallel group with $g5$, which however both describe the same regularity in the image. Therefore the smaller one is discarded as redundant. More precisely:

> Let $g_1||g_2$ and $g_a||g_b$ be two parallel groups. Then $g_1||g_2$ is defined to be *redundant* with respect to $g_a||g_b$ iff,
>
> $$\begin{aligned} GElem(g_1) &= GElem(g_a) \;\; \wedge \\ GElem(g_2) &\subset GElem(g_b) \end{aligned} \tag{3}$$

Currently, only parallel groups are implemented within the $2 \times 1$D -Level, since this is sufficient for our domain. With these steps all grouping hypotheses of the 1D and $2 \times 1$D level are available and together with the initial segmentation as primitive groups constitute the basis for the generation of closedness hypotheses.

### 4.3 2D level

A sequence of linear groups forms a hypothesis for a closedness, if their endpoints are hypothesized as pairwise close according to the proximity relation. An example are the

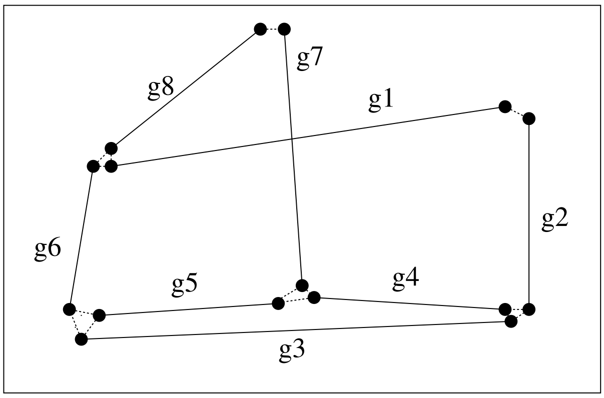



Figure 8: Example of a proximity graph: Endpoints of linear groups are shown as solid black circles with their position corresponding to their approximate location in the image. Contour edges are displayed with solid straight lines, regardless whether the underlying linear group is approximated by an elliptical arc or a straight line segment. Hypotheses for approximities are represented by proximity edges in the graph, shown as dotted lines. The simple cycle $g1, g2, g3, g6$ and $g1, g2, g4, g5, g6$ constitute hypotheses of closedness in this example, while $g1, g2, g4, g7, g8$ is not hypothesized due to the intersection of $g7$ and $g1$.

linear groups 0, 5, 4, and 1|2|3 in Figure 7(a). To detect such a closedness we define the *proximity graph*. As depicted in Figure 8, each endpoint of a linear group constitutes a node in this graph. Nodes are connected by two types of edges: Both endpoints of a linear group are connected by a *contour edge*, while endpoints of a proximity grouping are connected by a *proximity edge*. Each closedness is represented in this graph as a simple cycle with alternating types of graph edges. In addition, no intersections between linear groups and the gaps between endpoints (displayed as the proximity edges in the Figure 7(a)) are accepted for a valid hypothesis. We employ a simple graph search algorithm to find all such simple cycles to generate hypotheses of type closedness.

Again, we have to consider redundant hypotheses which may arise similar as for collinearities by “skipping” a linear group within a sequence using an appropriate proximity. This is shown in Figure 9(a) for an example. However, subfigure (b) demonstrates that it is not sufficient to consider only the sets of grouping elements to check for redundancy as is done for collinearities.

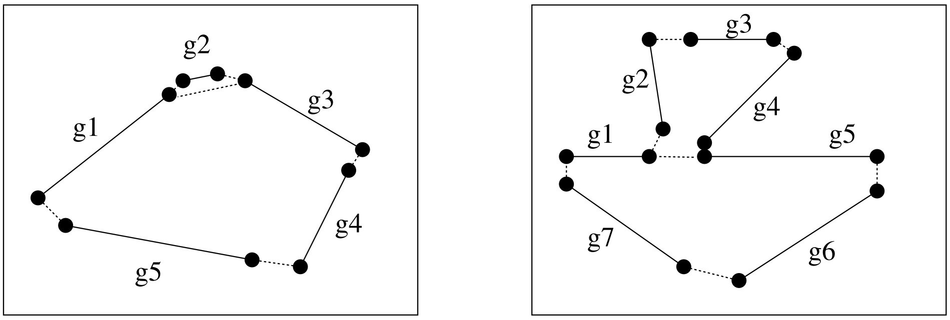
Therefore, given two closedness hypotheses $G_1$ and $G_2$, we define $G_1$ to be *redundant* with respect to $G_2$ if the following two conditions hold:

$$GElem(G_1) \subseteq GElem(G_2) \qquad \wedge \tag{4}$$

$$\frac{|F(G_1) \cap F(G_2)|}{max(|F(G_1)|, |F(G_2)|)} \geq 0.95, \tag{5}$$

where we denote with $|F(G_{1,2})|$ the area enclosed by $G_1$ and $G_2$.

The second condition enforces an almost complete overlap of two closedness hypotheses to be redundant, to exclude cases as shown in Figure 9(b). The choice of the threshold




(a) (b)

Figure 9: Skipping of short linear groups when building a closedness.
(a) Due to the proximity ($g1.g3$) the two redundant hypotheses of closedness $g1 \circ g2 \circ g3 \circ g4 \circ g5$ and $g1 \circ g3 \circ g4 \circ g5$ are hypothesized.
(b) In this example the hypotheses of closedness $g1 \circ g2 \circ g3 \circ g4 \circ g5 \circ g6 \circ g7$ and $g1 \circ g5 \circ g6 \circ g7$ do not represent redundant hypotheses.

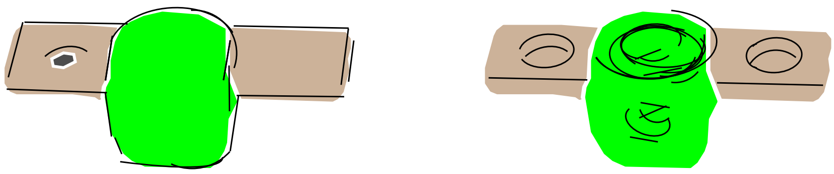
of 0.95 is not critical, it has only to account for small variations due to linear interpolation between neighboring endpoints within a closedness.

Additionally, if a closedness contains two linear groups being hypothesized as a parallel group, this is used as additional evidence for subsequent judgment. In the example of Figure 7 this is reflected in the only hypothesis for a closedness. With this step, the generation of hypotheses within the grouping hierarchy using only contour-based information is complete. Since these hypotheses are based only on local evidence they have to be judged in a more global context to yield a globally consistent interpretation of the image data. Before we describe the use of a Markov Random Field for this judgment in Section 6, we first consider cue integration of region segmentation to mutually enhance the quality of the grouping hypotheses and region segmentation.

# 5 Cue Integration

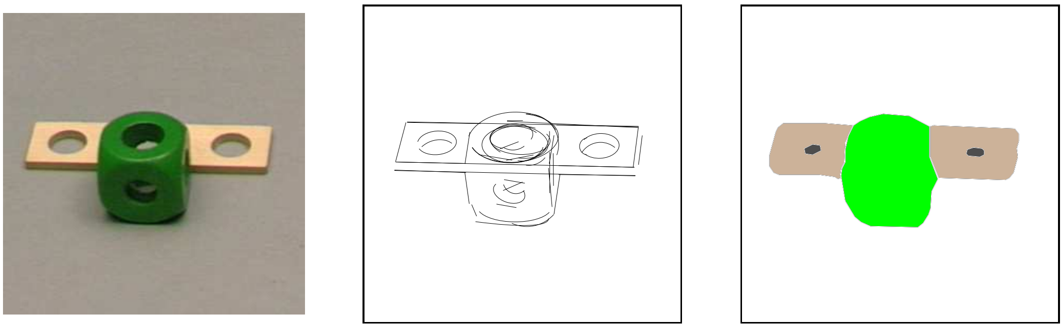
To integrate contour- and region-based information we propose to combine both cues at the level of contour groupings and complete regions, because the contours and their groupings contain more abstract information about the global structure of an image than pixel based edges do. To this end, we define *matches* between regions and linear groups. In our approach a match describes a relation between a detected region and a linear group in the sense of being caused by the same event in the scene, e.g. the projection of the same surface of an object. Matches are distinguished into *boundary* and *structural matches*. Boundary matches establish relations of groupings approximating the boundary of a region (see Figure 10(a)), whereas structural matches relate groupings based on the structure (or texture) of an object surface to the region approximating that surface (see Figure 10(b)).

These matches are employed to resolve the drawbacks inherent to both segmentation techniques by combining them. A typical drawback of region segmentation occurs in situations when a projection of an object surface is split into parts due to occlusion by



(a) Boundary match (b) Structural match

Figure 10: Different types of matches

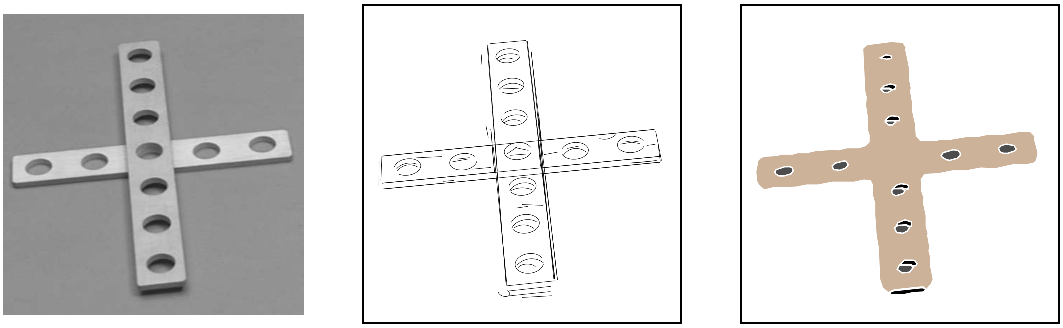


(a) (b) (c)

Figure 11: (a) Image with parts of BAUFIX parts (cube and wooden bar), (b) contour segmentation and linear groups, (c) region segmentation using pixel based color classification [20]

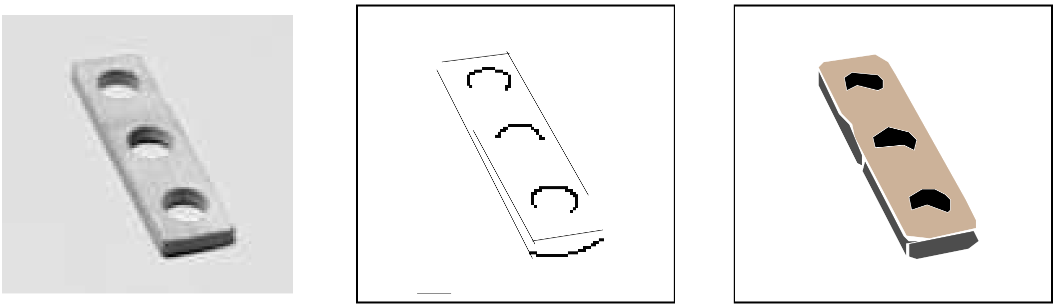
other objects: any region segmentation will produce disconnected regions for the bar in Figure 11. The contour-based groupings help to resolve this drawback by generating linear groups for the fragmented contours of the projected, partially occluded surface of the bar (see Figure 11(b)): We find a simple cycle of linear groups within the proximity graph for the correct surface projection of the bar and can use the information, that elements of this cycle match the two disconnected regions of the bar. This yields a cue to hypothesize an alternative region for the correct surface. Another situation occurs if two wooden bars overlap (see Figure 12). Region segmentation techniques inherently tend to generate one region for both bars, because their projections is of homogeneous color and the inhomogeneities at the boundary between both surfaces are usually insufficient for separating them. In this case the linear groups and detected simple cycles for the surfaces of the two bars support a separation of the region into the correct surface projections of the bars.

A characteristic disadvantage of contour segmentation is shown in Figure 13. In situations where the gradient magnitude is not sufficient to locally support a discontinuity in the image (see lower right corner of the bar in Figure 13(b)) the segmentation process fails. However, the region segmentation detects the boundary of the surface of the bar



(a) (b) (c)

Figure 12: (a) Image of overlapping wooden bars, (b) contour segmentation and linear groups, (c) region segmentation using pixel based color classification



(a) (b) (c)

Figure 13: (a) Image of a wooden bar, (b) contour segmentation and linear groups, (c) region segmentation using IslandGrowing [34]

(Figure 13(c)). This is due to fact, that region segmentation tries to detect homogeneous areas using more global information to segment the image, whereas the contour segmentation uses only local image data to detect discontinuities. In this situation the region boundary can help to close a path of linear groups in the proximity graph traversing the boundary to find the next linear group near the region boundary. See Figure 14 for a similar example, where region information can be used to close a path between $C_2$ and $C_3$.

These two examples show how we use the combination of region and contour segmentation including linear groups to support and constrain each other. Our focus is on the generation of hypotheses within the grouping hierarchy and will be described in detail in Section 5.2. Beforehand, the generation of matches and their classification into boundary and structural matches is described in the next subsection.

## 5.1 Matching linear groups to regions

Despite of the fact that both segmentation schemes describe the same events of the scene, they represent results differently. For their combination we use an uniform polygonal representation giving efficient means to calculate various measures. In the rest of this

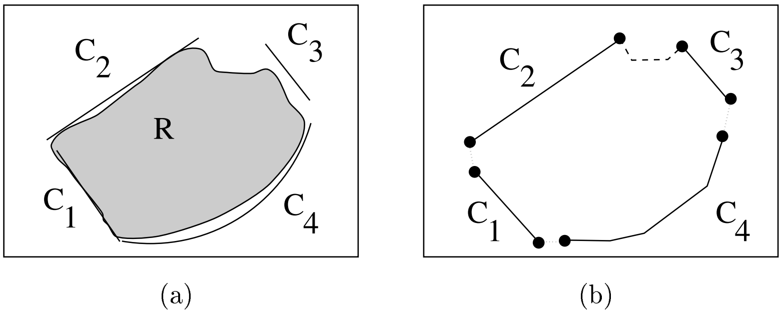



Figure 14: (a) Synthetic region and contour segmentation; (b) Sample proximity graph: nodes are displayed as black circles, contour edges corresponding to contour segments with solid lines; for $C_4$ the polygonal approximation used to check against intersection is depicted; proximity edges from contour based grouping are shown with dotted lines, while the proximity edge connecting $C_2$ and $C_3$ derived from region information is shown as a dashed line.

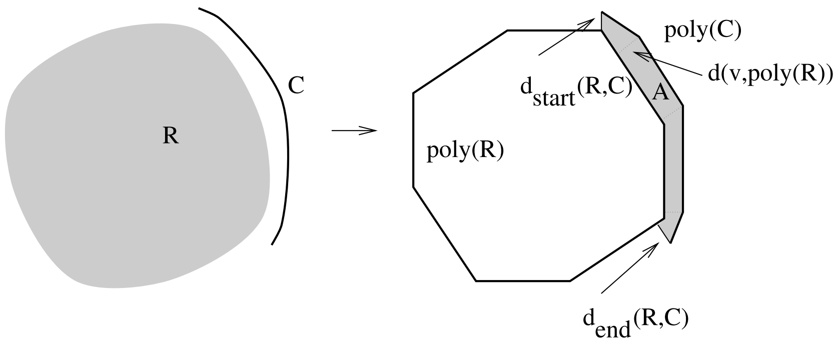



Figure 15: Matching of contour $C$ to region $R$

section the linear groups are denoted with $C$ and represented by open polygons $poly(C)$ approximating the underlying parametric models. The regions are represented by closed polygons $poly(R)$ for each region $R$, approximating the region boundary. In the following we match each linear group to a region if both probably belong to the projection of the same 3D-surface. Therefore each linear group is tested against all regions intersecting a neighborhood around the linear group. The test for a match is controlled by three measures, illustrated in Figure 15. The first two measures are the euclidean distances $d_{start}(R, C)$ and $d_{end}(R, C)$ of the endpoints of the linear group to the region boundary. To compute an estimate of the average distance between the linear group and the region boundary, the polygon of the linear group $poly(C)$ is projected onto the polygon of the region $poly(R)$. The area $A$ between the projection and the contour polygon is calculated. The third measure is then defined by: $d_{avg}(R, C) = \frac{A}{length(C)}$, where $length(C)$ is the length of the contour segment. This gives an estimate for the average distance between both polygons. These measures are used by a polynomial classifier of degree 5 to classify a match into boundary or structural match. The classifier is trained with a handlabeled training set of images of our domain. Additional constrains must be fulfilled to avoid incorrect matches based on configurations, which are not included in the handlabeled training set: a contour-based grouping must be inside the region half of its length to yield to a structural match and a linear group with small endpoint distance and big average distance, but outside the region, is not able to establish a boundary match with the region.

### 5.2 Using matches to detect simple cycles of contours

We now describe how these matches can be used to generate cycles of linear groups even if the proximity information of the grouping process is not sufficient.

The search for simple cycles of linear groups may fail in situations, where a proximity relation between two linear groups is missing to close a path. These situations can be resolved using the information which linear group matches a region with a boundary match: In Figure 14, the path $C_1 - C_2$ can not be closed to $C_3 - C_4 - C_1$, due to the missing contour-based proximity information between $C_2$ and $C_3$. However, we can use the matching of $C_2$ and $C_3$ to the same region $R$ to complete the path along the region boundary. We call this a *region based proximity* between the endpoints of $C_2$ and $C_3$ and add these to the proximity graph, which is used to find closedness hypotheses as defined in Section 4.3. The following describes this idea in detail.

For this process only boundary matches between linear groups $C$ and regions $R$ are considered. For each region all linear groups, related to the region through a boundary match, are linearly ordered around the region. This order is defined by the projected start- and endpoints of the linear group onto the region polygon. Thus, for the endpoint of one matching linear group, the start of the next matching linear group $\tilde{C}$ can be found along the region boundary of the same matched region $R$. Adding region based proximities to the proximity graph for perceptual grouping of contour segments is controlled by the distance of the endpoints of $C$ and $\tilde{C}$. Only if the distance between the endpoints of $C$ and $\tilde{C}$ along the region boundary of $R$ is smaller than the length of each segment and there is not already a proximity edge in the graph between them, a new region based proximity edge is inserted.

These edges are used in the same way as proximity edges are used in the proximity graph to detect simple cycles as described in Section 4.3 Thus, region information is employed to enhance the quality of contour-based interpretation of the image data.

## 6 Global Judgment using Markov Random Fields

The generation of grouping hypotheses is complete with this incorporation of information from region segmentation yielding additional closedness hypotheses. To judge grouping hypotheses in a global context two main requirements have to be satisfied:

- The interaction of the locally generated grouping hypotheses has to be modeled, which may be competing or supporting.
- The global optimal state has to be determined by an appropriate control algorithm.

We use a Markov Random Field (MRF) to find the global optimal state for the locally interacting grouping hypotheses. This process takes advantage of the neighborhood system for the formation of local interacting grouping hypotheses. In the following subsection we give a brief introduction to MRFs and the notation we use in the rest of the paper. For a thorough treatment of MRFs, especially applied to computer vision see [11, 23]. In the subsequent subsections we describe our use of MRFs for perceptual grouping.

## 6.1 Markov Random Fields

To represent a given problem with a MRF, we define a set of entities or *sites* $\mathcal{S} = \{s_1, \ldots, s_N\}$. In image analysis sites typically correspond to pixels or image primitives like regions. Each site has to be assigned a *label* drawn from the label set $\mathcal{L} = \{l_1, \ldots, l_Q\}$ describing the state or interpretation of the site. To this end, a random variable $X_i$ is associated to each site which accordingly takes values from the label set $\mathcal{L}$. The set of all *configurations* is denoted with $\Omega = \{\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N) | \omega_i \in \mathcal{L}\}$ where $\omega_i$ is the value of $X_i$. Interactions between sites result in a neighborhood system which may equivalently be expressed as an undirected graph with the sites as nodes.

The distribution $P(\boldsymbol{\omega})$ of Markov Random Fields is characterized by a local property stating that the value of a random variable $X_i$ is only dependent on the value of its neighbors in the graph. A more practical approach to model this Markovian property is given by the Hammersley-Clifford theorem (see [3]), which states its equivalence to a Gibbs distribution. A Gibbs distribution is given as

$$P(\omega) = \frac{1}{Z} e^{\frac{-U(\omega)}{T}}, \tag{6}$$

where $Z$ is a normalizing constant and $T$ represents the temperature, which controls the flatness of the distribution. The energy function $U(\omega)$ is defined as

$$U(\omega) = \sum_{\mathbf{c} \in \mathcal{C}} V_{\mathbf{c}}(\omega), \tag{7}$$

where $\mathcal{C}$ is the set of cliques and $V_{\mathbf{c}}$ are the associated clique potentials. As usually, a clique is a subset of nodes (or sites) which are pairwise neighbors within the graph.

However, for image interpretation one is interested in the a posteriori probability $P(\omega|\mathbf{d})$ given the image data $\mathbf{d}$ as observations. In the following we assume conditional independence between the observations $d_i$ at each site which is not exactly true but common practice in MRF approaches to computer vision (cf. [6]). Under this assumption and with Bayes's rule we can derive the a posteriori energy as:

$$U(\omega|\mathbf{d}) = \sum_{\mathbf{c} \in \mathcal{C}} V_{\mathbf{c}}(\omega) - T \sum_{i=1,\ldots,N} \log P(d_i|\omega) \tag{8}$$

Minimizing this a posteriori energy results in a maximum a posteriori estimate of the random field and gives a globally consistent interpretation of the (image) data.

## 6.2 Perceptual Grouping with a MRF

In our approach each grouping hypothesis is modeled as one site of the graph underlying the MRF, which we call *grouping graph*. As a consequence, in contrast to most other approaches using MRFs, different sites may interpret a common subset of the image data. (An exception is the system described in [7] to detect subjective contours.) Referring to the example in Figure 7 (see Page 13), a given set of edge elements is approximated by the contour segment 1 which is part of the hypothesis for the collinear group $1|2|3$. This in turn is hypothesized as an element of the closedness $0 \circ (1|2|3) \| 5 \circ 4$. Thus, those three sites of the MRF all are related to a common set of pixels. The same may be true for competing hypotheses, which are discussed below.

The random variables associated to each node represent the significance of the hypothesis being a correct interpretation of the image data similar to the approach in [19]. It has to reflect as well the evidence given by the image data as the prior knowledge about context dependencies including Gestalt principles. The significance takes discrete values in the interval $[0\,;1]$, where large values encode high significance of a hypothesis. Thus, the set of labels is given as

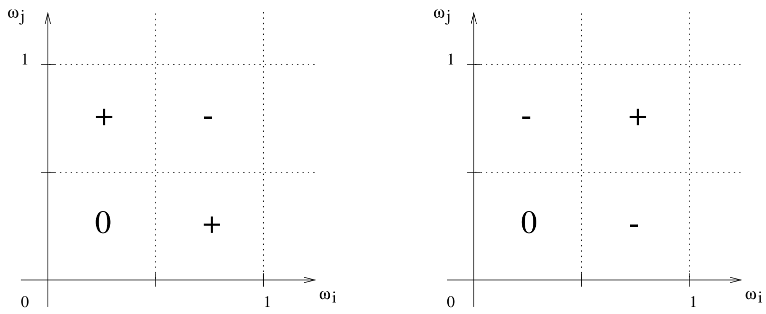
$$\mathcal{L} := \{0, \triangle_{\text{sig}}, 2\,\triangle_{\text{sig}}, \ldots, 1\}\,, \tag{9}$$

where we use $\triangle_{\text{sig}} = 0.05$ as discretization steps for all our experiments.

Again, in contrast to other approaches, the neighborhood system of the graph does not represent spatial neighborhood between grouping hypotheses, but rather models the dependencies between hypotheses with regard to a globally consistent interpretation of the image data. As mentioned in Section 3 it is constructed with supporting and competing undirected edges: a hypothesis is supported by all its parts, due to the fact, that a correct grouping hypothesis requires its parts as a prerequisite. On the other hand the hypothesis supports its parts, because it confirms them being a correct interpretation of the image data. In contrast, competing edges are defined between hypotheses, which model contradicting interpretations of the image data: Two collinear or two curvilinear groupings compete, if they share common contour segments. Two parallel groups or two closedness hypotheses compete with each other, if they again share a common part and in addition overlap a common region of the image. Thus, supporting relations occur mainly between the different levels of the hierarchy, while competing relations connect nodes within one level. An example for a complete grouping graph is given in Figure 7(b). To complete the definition of the MRF the clique potentials $V_{\mathbf{c}}$ are developed in the rest of this section.

## 6.3 A prior knowledge

Potentials of single-site cliques, i.e. cliques with one element, describe the prior likelihood of a label at a given site without image information. Since we can only assume uniform distribution for the significance we neglect single-site clique potentials for the prior energy. The pair-site cliques formulate the relations of support and competition between two given nodes as described in the last subsection. For supporting hypotheses the potential is designed as follows (see Figure 16 for illustration): Two significant




(a) (b)

Figure 16: Illustration for the qualitative design of the potential for pair-site cliques: (a) support and (b) competition.

grouping hypotheses are consistent with each other yielding a negative clique potential. A significant hypothesis and a hypothesis with low significance in contrast are contradicting and hence the assigned clique potential is positive. Two hypotheses with low significance can neither be interpreted as contradicting nor consistent resulting in a neutral clique potential of zero. This qualitative definition is implemented with a Gaussian

$$\psi(\omega_i, \omega_k) = \lambda \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\left((\omega_i-\mu_i)^2+(\omega_k-\mu_k)^2\right)}, \tag{10}$$

where $\sigma = 0.2$ and $\lambda = 0.25$ yield smooth transitions between the different cases:

$$V_{\mathbf{c}}(\omega_i, \omega_k) = \begin{cases} -\psi(\omega_i, \omega_k) & : 0.5 < \omega_i, \omega_k \leq 1 \\ & \phantom{:} \text{ with } \mu_i = 1,\ \mu_k = 1 \\ \psi(\omega_i, \omega_k) & : 0 \leq \omega_i \leq 0.5 < \omega_k \leq 1 \\ & \phantom{:} \text{ with } \mu_i = 0,\ \mu_k = 1 \\ \psi(\omega_i, \omega_k) & : 0 \leq \omega_k \leq 0.5 < \omega_i \leq 1 \\ & \phantom{:} \text{ with } \mu_i = 1,\ \mu_k = 0 \\ 0 & : 0 \leq \omega_i, \omega_k \leq 0.5 \end{cases} \tag{11}$$

In addition the supporting clique potential is normalized by the number of parts of a hypothesis. This is necessary, because for two competing hypotheses the one with more parts would receive more support just because of the larger number of supporting edges within the MRF graph. Therefore, the potential for supporting relations is defined as:

$$V_{\mathbf{c}}^{support}(\omega_i, \omega_k) = \frac{1}{|GElem(h_i)|} V_{\mathbf{c}}(\omega_i, \omega_k), \tag{12}$$

where $h_i$ is the grouping hypothesis represented by site $s_i$ including $h_k$ (associated to $s_k$) as a grouping element.

The potential for competing relations is defined the other way round, since two competing hypotheses both being significant constitute an inconsistent interpretation. This is modeled by a positive contribution to the energy. In contrast, a significant hypothesis and a hypothesis with low significance are compatible and result in a negative potential:

$$V_{\mathbf{c}}^{compete}(\omega_i, \omega_k) = -V_{\mathbf{c}}(\omega_i, \omega_k) \tag{13}$$

According to the definition of the grouping graph, also cliques with more than two elements do occur. These are built either by two conflicting groupings and a common grouping element (in the example of Figure 7(a) the parallel groups $(1|2|3)\|8$ and $(1|2|3)\|5$ with their common part $1|2|3$) or by a set of pairwise conflicting hypotheses (the collinear groups $1|2$, $1|3$, and $1|2|3$ in the example). Currently, we do not consider these cliques explicitly, but rather model this context dependency implicitly using pair-site cliques.

## 6.4 Data dependency

With these clique potentials the model of the prior knowledge is concluded. To complete the definition of the posteriori energy $U(\omega|\mathbf{d})$ we have to define the terms $T \log P(d_i|\omega)$ of equation (8), where we assume independency of the observations at different sites as mentioned. These can be viewed as single-site clique potentials of the a posteriori MRF defined by $U(\omega|\mathbf{d})$ and they are modeling the data dependency in our approach.

To this end a data driven significance $data\_sig(s_i)$ is defined for each hypotheses. It defines a measure how well the hypothesis conforms to the underlying grouping principle when interpreting local image data by the hypothesized grouping. For generation of hypotheses as described in Section 4, we employed a hard decision regarding this compliance with Gestalt laws, while we use it in a quantitative fashion for global judgment of hypotheses. The data driven significance again maps to the interval $[0\ ;1]$ and is described below for the different types of groups in detail. The clique potential measures the deviation of this data driven significance to the significance within the global context as assigned by the label $\omega_i$ of the current configuration:

$$V^{data}_{\{s_i\}}(\omega_i) := T \log P(d_i|\omega) := |\omega_i - data\_sig(s_i)| \tag{14}$$

As stated above, the data driven significance captures again compliance with the underlying grouping principle and follows the same considerations as employed for generation as described in Section 4.2. For a collinearity at site $s_i$ the first requirement of forming a straight line is now formalized by the mean approximation error $\bar{d}(s_i)$ normalized to $[0\ ;1]$ using minimum and maximum approximation errors of all collinear groups:

$$data\_approx(s_i) := 1 - \frac{\bar{d}(s_i) - \min\limits_{s_j} \bar{d}(s_j)}{\max\limits_{s_j} \bar{d}(s_j) - \min\limits_{s_j} \bar{d}(s_j)} \tag{15}$$

The second condition for small gaps between contour segments is quantified with the mean gap length $\bar{g}(s_i)$ and the mean segment length $\bar{l}(s_i)$ of all grouping elements as

$$data\_gap(s_i) := \max\left\{ \big(\bar{l}(s_i) - \bar{g}(s_i)\big)/\bar{l}(s_i), 0 \right\}, \tag{16}$$

and the data driven significance $data\_sig(s_i)$ is given as the arithmetic mean of both, $data\_approx(s_i)$ and $data\_gap(s_i)$. For curvilinear groups the definition is exactly analogous, while for proximities only $data\_gap$ is needed.

For a parallel group $s_i$ with grouping elements $g_1$ and $g_2$ the admissible interval of orientation difference is mapped to the interval [0.5 ; 1]:

$$data\_par(s_i) := 1 - \frac{\triangle(g_1, g_2)}{2 \cdot \Delta_{\mathrm{par}}}. \tag{17}$$

In contrast to the generation process the overlap $\tau(s_i)$ is computed exactly, projecting one contour segment onto the other and normalized with respect to mean length $\bar{l}(s_i)$ of $g_1$ and $g_2$:

$$data\_overlap(s_i) := \tau(s_i)/\bar{l}(s_i) \tag{18}$$

The third component of spatial closeness is make concrete using the mean distance $\kappa(s_i)$ between $g_1$ and $g_2$, which again is normalized to [0 ; 1] using the median $\bar{\kappa}$ of distances from the hand labeled training set: If $\kappa(s_i)$ coincides with $\bar{\kappa}$ the maximum value of 1 results. For smaller $\kappa(s_i)$ it decreases linearly, otherwise hyperbolically:

$$data\_dist(s_i) := \begin{cases} \kappa(s_i)/\bar{\kappa} & \text{for } \kappa(s_i) \leq \bar{\kappa} \\ \bar{\kappa}/\kappa(s_i) & \text{otherwise} \end{cases} \tag{19}$$

These three values are linearly combined as:

$$data\_sig(s_i) := 0.2\ data\_par(s_i) + 0.6\ data\_overlap(s_i) + 0.2\ data\_dist(s_i) \tag{20}$$

The weights have been derived from initial experiments on a small set of images.

For closedness hypotheses the only requirement in the current system is the need for small gaps, which was enforced be the need for proximities between endpoints in the proximity graph of Subsection 4.3. Now it is modeled with a data driven significance equal to *data_gap* as defined in equation (16). This definition is a pure domain independent one and therefore no bias towards particular shapes is introduced. If desired, however, this could be accomplished easily with an appropriately extended definition of the data driven significance for closedness hypotheses.

## 6.5 Energy Minimization

Given this definition of the clique potentials the a posteriori energy is summarized as:

$$\begin{aligned} U(\omega|\mathbf{d}) = & \sum_{\{s_i,s_k\} \in \mathcal{C}_2^{support}} V_{\{s_i,s_k\}}^{support}(\omega_i, \omega_k) + \\ & \sum_{\{s_i,s_j\} \in \mathcal{C}_2^{compete}} V_{\{s_i,s_j\}}^{compete}(\omega_i, \omega_k) + \sum_{s_i} V_{\{s_i\}}^{data}(\omega_i) \end{aligned}$$

This shows, that all clique potentials contribute independently of the level within the hierarchy and the spatial extent of grouping hypotheses. This models scale independence of grouping phenomena and gives the different grouping principles employed at various levels of the hierarchy the same influence.

The overall energy minimization is controlled by “Highest Confidence First” according to [6]. This method is deterministic, fast, independent of a specific initialization, and is shown to yield a local minimum which compares well to other minimization procedures.

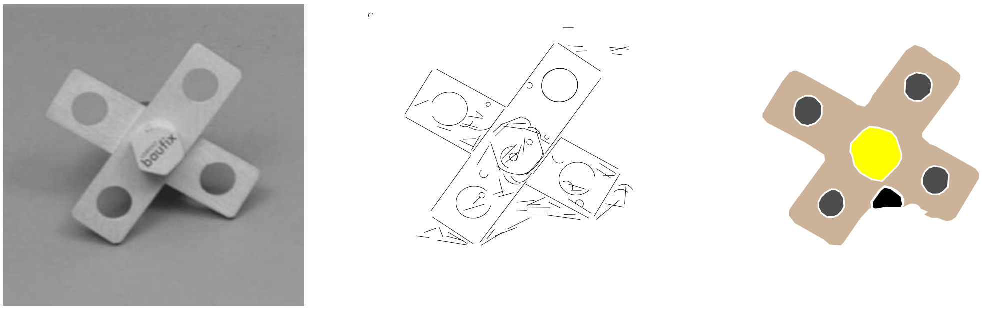
# 7 Results

In this section we present results of perceptual grouping and cue integration for two real images of our domain (see Figures 17, 20). To this end the results will be discussed for the different types of grouping successively. With the description of generated hypotheses, their judgment in the Markov Random Field will be described simultaneously. Concerning the judgment, in the following a hypothesis with label $\omega_i \geq 0.6$ will be called *significant* and with $\omega_i < 0.6$ *insignificant*. At the end of the section the combination of contours and regions will be discussed, demonstrating its benefits generating hypotheses of closedness using region-based proximities in the proximity graph.

The aim of our approach is to yield a global contour-based description of implicit structure in the image. An initial contour segmentation represents the projection of visible boundaries of object surfaces we are interested in and additional texture. Figure 17(b), 20(b) show these initial contour segments described by parametric models of elliptical arcs and straight lines. However, these contour segments may be fragmented due to low contrast. An example for this can be seen in Figure 17(b): a gap between two short line segments on the left side of the bolt head is due to low contrast. On the other hand, fragmentation can be caused by occluding objects. In Figure 17(b) the top bar occludes the lower one resulting in fragmentation of the line segments describing the boundary of the lower bar.

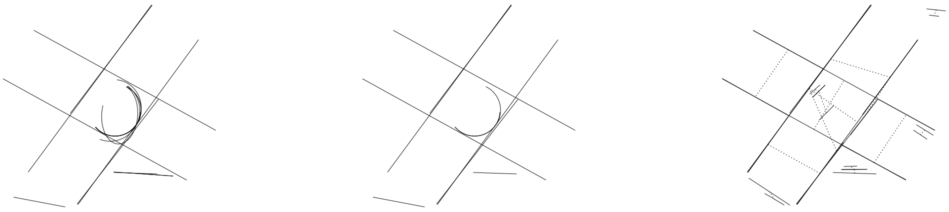
The good initial contour segmentation yields only a few fragmented line segments. In both images this fragmentation is due to occlusion by other objects: in Figure 18(a) boundaries of both bars are fragmented as in Figure 21(a) for the bars forming the wing, tail and body of the glider. But, as shown in Figures 17(b), 20(b) these fragmentations of boundary contours could be resolved by perceptual grouping. In addition, some of the hypothesized collinearities yield an incorrect interpretation of the image data: in Figure 21(a) collinearities group line segments of different bars. On the other hand, one correct grouping on the right side of the bar forming the body of the glider is labeled insignificant. This is due to the fact, that the closedness, this collinearity is part of, is labeled insignificant (see Figure 22(b)). All other correct hypotheses are labeled significant in a global context.

The grouping of elliptical contour segments into curvilinearities is not as successful as the collinear grouping process. This is due to difficulties in estimating the orientation of tangents at the endpoints of the elliptical approximations. Examples can be seen in Figure 18(a): Curvilinearities are generated for the elliptical segments of the bolt, and these in turn are grouped with other elliptical arcs of its shadow on the bar. Similar situations occur in the image of the glider yielding incorrect groupings for the screws fixing the bars of the tail unit and the wing. In detail, an example occurs at the wing: elliptical arcs of different screws are grouped as curvilinear (see Figure 21(a)). However, all these incorrect groupings are labeled insignificant by the Markov Random Field because they are not part of higher level grouping hypotheses and have low data driven significance (see Figure 21(b), 18(b)). On the other hand, parts of the boundary of the bolt head in the image in Figure 17 are partially approximated with elliptical arcs and three of these are grouped to form one curvilinearity yielding a correct grouping (see Figure 18(b)).



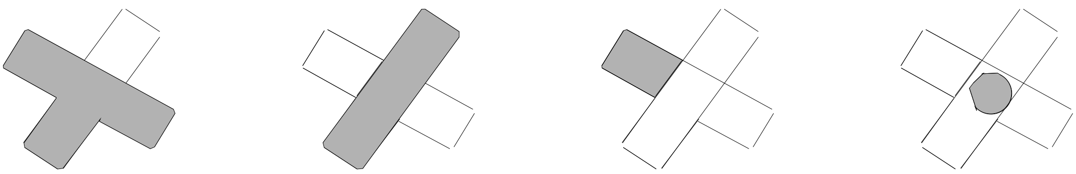

(a) (b) (c)

Figure 17: (a) Image of wooden bars forming a cross, (b) initial contour segmentation, (c) region segmentation



(a) (b) (c)

Figure 18: Grouping hypotheses of level 1D and 2×1D: (a) all generated collinearities and curvilinearities, (b) significant collinear and curvilinear groups, (c) significant parallel groups (parallel relation between linear groups depicted as doted line)



(a) (b) (c) (d)

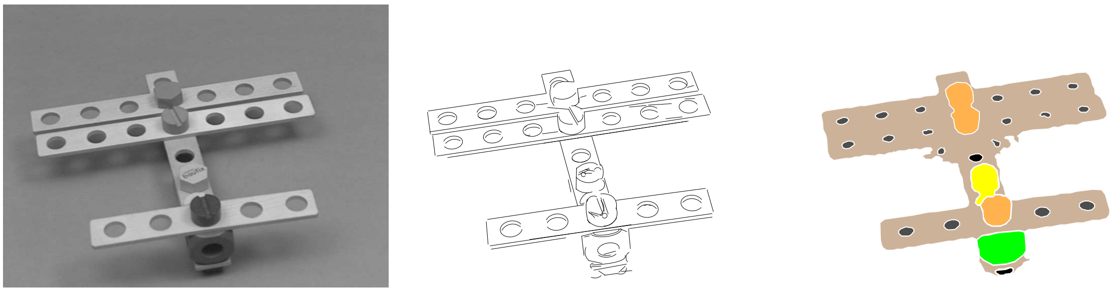
Figure 19: (a)-(c) Significant closedness hypotheses without using region information, (d) additional significant closedness using region based proximities

In Figure 21(b) similar examples are depicted: for the screws fixing wing and tail unit correct groupings are generated and all of them are labeled significant. Summarizing, grouping of contour segments into collinearities and curvilinearities successfully employs the Gestalt principle of good continuation to capture significant image structure in the presence of occlusion. (Note, that no primitive groups but only groupings are depicted in Figure 18(b), 21(b).) A detailed discussion of the proximity groups is difficult due to their number. Therefore their quality will be shown by closedness groupings as proximities are used to generate closedness hypotheses in the proximity graph.

Parallel groups are generated for linear groups describing the boundary of the longer side of the bars and also for some parallel line segments resulting from noise and shadow in the images (see Figure 18(c), 21(c)). In addition, the initially fragmented line segments in both images are hypothesized to be parallel. In detail, for the occluded bar in the image of Figure 17 three parallel groups are generated, each one for the fragmented line segments outside the occluded area and one for the collinearities which group these fragments. Similar groups are generated for the other bar and for the wing, tail and body of the glider in the image in Figure 20. The large number of significant parallel groups for both images is due to supporting relations of collinearities and part-of relations in most of the closedness hypotheses in the MRF. Based on the hierarchy these groupings receive support from higher and lower level groupings yielding global structure information concerning parallelism in the images.

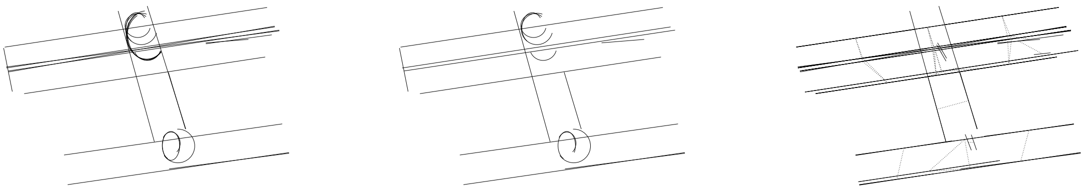
As the last grouping type, we discuss closedness groupings. In Figures 19, 22 subsets of all hypothesized closedness groupings are depicted. In the image of the glider for each of the bars forming the wing of the glider, five closedness hypotheses are generated. Since these hypotheses interpret the same structure in the image, they are competing with each other in the MRF. Only two hypotheses of each bar are labeled significant (see Figure 22(a)). Additionally, a closedness is generated for the tail unit and for the tail including the body of the glider (see Figure 22(a),(b)). However, only the closedness for the tail unit is labeled significant, because both have a competing relation to each other. This is due to the fact, that both closedness hypotheses have common hypotheses of collinearities at the tail unit and therefore the energy minimization process tends to decrease the label for one hypothesis. In this case the one for the tail including the body of the glider is labeled insignificant. This situation can only be resolved by incorporating domain dependent knowledge into the data driven significance as already mentioned in Section 6. Similar situations occur in the image depicted in Figure 17: the graph search for simple cycles results in five closedness hypotheses. Two of them merge the surface of one bar with one part of the other bar as depicted in Figure 19(a). Two other closedness hypotheses describe the correct surface projection of the bars as in Figure 19(b), and the last one is depicted in Figure 19(c). Due to this constellation the closedness for the correct surface projection of the lower bar, and one of the merging closedness hypotheses are labeled insignificant in the Markov Random Field. This is for the same reason as described in connection with the glider. Generally it can be stated, that this type of misleading interpretation of the structure occurs in situations, where many groupings interpret the same underlying data without domain dependent knowledge.

These examples show the powerful capabilities of Gestalt principle implemented for



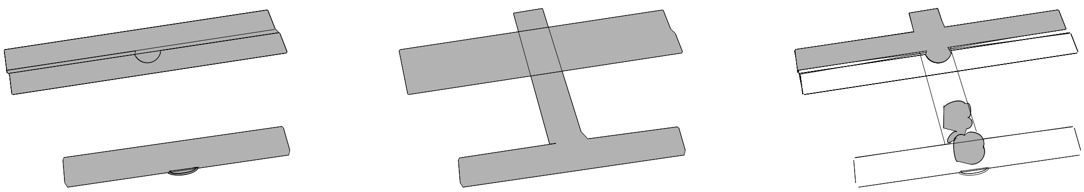
(a) (b) (c)

Figure 20: (a) Image of BAUFIX parts forming a glider, (b) initial contour segmentation, (c) region segmentation



(a) (b) (c)

Figure 21: Grouping hypotheses of level 1D and 2×1D: (a) all generated collinearities and curvilinearities, (b) significant collinear and curvilinear groups, (c) significant parallel groups (parallel relation between linear groups depicted as doted line)



(a) (b) (c)

Figure 22: (a) Significant closedness hypotheses without using region information, (b) insignificant closedness hypotheses (the one for the whole wing uses region based proximities) (c) additional closedness using region-based proximities with $\omega_i \geq 0.5$

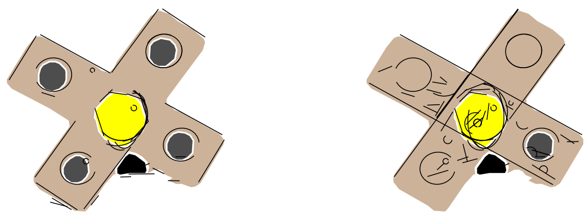
the generation of grouping hypotheses. The judgment in a global context yields a selection of the relevant hypotheses to extract relevant structures in the image. In Table 2 the number of generated and significant hypotheses for the images in Figures 17 and 20 are listed. The mean values for the corresponding numbers of a larger test set of 10 images are listed, too.

| | initial segments | | | | grouping hypotheses | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | line | | ellipse | | collinear | | curvilinear | | parallel | | closedness | |
| | hyp | sig | hyp | sig | hyp | sig | hyp | sig | hyp | sig | hyp | sig |
| cross | 78 | 50 | 29 | 9 | 10 | 8 | 6 | 2 | 23 | 15 | 5 | 3 |
| glider | 73 | 65 | 71 | 37 | 15 | 12 | 12 | 6 | 48 | 29 | 15 | 7 |
| mean of 10 images | 105 | 51 | 50 | 14 | 12 | 8 | 5 | 3 | 44 | 27 | 8 | 5 |

Table 2: Number of hypothesized (hyp) and significant groupings (sig) with $\omega_i \geq 0.6$ for the cross in Figure 17 and the glider in Figure 20. In the third row the mean for a set of 10 images of different complexity is listed.

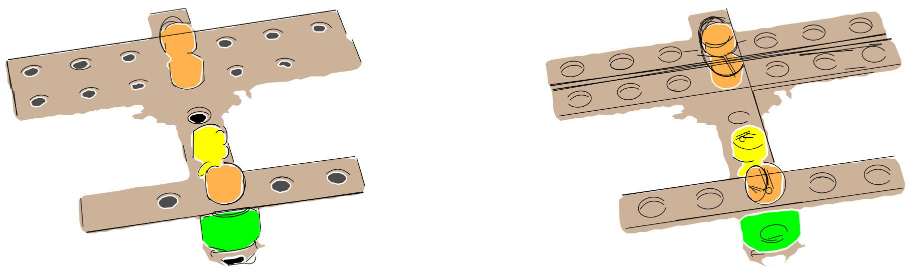
In some situations the fragmentation of contours could not be resolved as described in connection with the proximity grouping at the bolt head in Figure 17(b). This hinders the strict contour-based grouping to extract all structures in the image data. For this reason we combine the contour-based approach with region segmentation. Region segmentation is realized with an pixel based polynomial classifier of degree six, labeling each pixel with one of ten colors of the objects in our domain. Smoothing this label image yields color based regions represented and approximated by closed polygons [20]. The results of region segmentation for both images are depicted in Figure 17(c), 20(c). However, for the combination we propose, any region segmentation scheme can be used instead. To match regions and linear groups we again use a polynomial classifier for the combination as described in Section 5.1. For our domain this classifier was trained on a set of eleven segmented images. In Figures 23(a),(b) and 24(a),(b) the results of the combination are depicted: 42 boundary and 64 structural matches for the image in Figure 17 and 92 boundary and 106 structural matches for the image of the glider are generated. It can be seen that most of the matches are correct. In case of the occluded bar in Figure 23(a) the contours for the bottom boundary are not labeled as boundary matches, because they are inside the region and too far away from the boundary. On the other hand most of the collinearities for the boundaries of the bars are combined as structural match for the underlying region yielding a description of the structure inside the region (see Figure 23(b)). Similar results can be seen for the glider in Figure 24. The missing structural match for the left collinearity describing the boundary of the body of the glider in Figure 24(b) is due to a misclassification of the polynomial classifier and according to conditions as described in Section 5.1.

With these matches we enhance the generation of closedness hypotheses using the boundary matches to generate region-based proximities in the proximity graph (see Section 5.2). In detail, in Figure 17(b), on the left side of the bolt head, a gap between two short line segments approximating the boundary is too wide due to low contrast in the



(a) Boundary match (b) Structural match

Figure 23: Matching of linear groups and regions



(a) Boundary match (b) Structural match

Figure 24: Matching of linear groups and regions

| | Contour | Grouping | | | MRF |
|---|---|---|---|---|---|
| | initial | 1D and | 2D | | |
| | segmentation | $2 \times 1D$ | contour-based | contour+region-based | labeling |
| cross | 2814 ms | 432 ms | 597 ms | 932 ms | 116 ms |
| glider | 12962 ms | 832 ms | 2030 ms | 4700 ms | 449 ms |
| mean of 10 images | 2540 ms | | | | 600 ms |

Table 3: Runtime for the images in Figure 17 and Figure 20 and the mean for 10 images of our domain running on DEC Alpha 500/400MHz

image data. Consequently it is not possible to establish a contour-based proximity group between them. In Figure 19(d) a closed cycle for the bolt head is depicted, which is hypothesized using the region-based proximity between the two short line segments matched as boundary matches to the region of the bolt head. Similar examples are shown in Figure 22(b),(c). The closedness for the whole wing in Figure 22(b) is labeled insignificant due to competing relations with the closedness hypotheses for the bars. However, in Figure 22(c) additional cycles are found for the screws at the tail of the glider which are labeled with $\omega_i \geq 0.5$. The generation of alternative regions as described in Section 5 is under way to be implemented.

In Table 3 the runtime for initial contour segmentation, generation of hypotheses, additional matching, and the judgment in the Markov Random Field for the discussed images are listed. It can be seen that an additional matching of contours results in a small overhead in runtime, whereas it significantly enhances the generation of closedness hypotheses.

# 8 Conclusion

To integrate principles of perceptual organization as proposed by the Gestalt psychologists into a contour-based approach to computer vision, we introduced a grouping hierarchy with increasing complexity of grouping types. Contour segments from an initial segmentation form the basis of this hierarchy. The first, 1D-Level employs the Gestalt principles of good continuation and proximity to organize contour segments into collinearities, curvilinearities, and proximities. These in turn are grouped to symmetric and parallel groups in the successive $1 \times 2$D-Level. The currently most complex 2D-Level realizes the principle of closure to form closedness groups from groupings of the lower levels. Currently, we use contour segments approximated as straight line segments or elliptical arcs, extended elliptical arcs as curvilinearities, and only parallel groups within the $1 \times 2$D-Level as appropriate for our domain. However, the design of the hierarchy is not limited to these choices. Additional types of parametric contour models or non-parametric representations of contour segments may be integrated into the model, if needed. The same is true for adding further levels to the hierarchy to represent even more complex grouping types. As an example, it is fairly straightforward to add another level to represent collinear or curvilinear groups of closed contours again making use of the principle of good continuation.

Perceptual organization of contour segments within this hierarchy is divided into two stages. First, grouping hypotheses are generated bottom up with respect to the hierarchy. To implement the various Gestalt principles for the 1D- and $1 \times 2$D-Level, we introduce the concept of areas of perceptual attentiveness to model the spatial extent of grouping effects playing an important role not only for proximity groupings. These search areas are applied in conjunction with conditions on orientation difference, gap length, and overlap of primitives to be grouped. The shape and size of the areas of perceptual attentiveness as well as the necessary thresholds for the conditions are derived from a hand labeled training set of images. To generate groupings of the 2D-Level, we define the proximity graph representing all hypotheses of the 1D-Level. In this graph we detect simple, non-intersecting cycles as closedness hypotheses. If such a cycle in addition contains linear groups hypothesized as being parallel, this is used as additional evidence, concluding the first stage.

Before judging these hypotheses in a global context, cues from region segmentation are introduced into this contour-based process. To do this, we define a linear group to match a region if both are related to a common surface projection. Matches are distinguished as boundary or structural, depending on the contour segment to belong to the boundary or texture of the region. A polynomial classifier is used to implement this matching process. The resulting boundary matches are used to augment the proximity graph with additional edges, if the matching process gives region-based evidence for contour segments to delineate a common region. Work is under way, to inversely enhance the result of region segmentation using the matches. This aims at generating alternative regions making use of contour information and particularly of the principle of good continuation made explicit in the groupings of the 1D-Level.

While this completes the generation of hypotheses, these processes are still based on rather local evidence. To judge hypotheses in a more global context a Markov Random Field is used. Each grouping hypothesis is modeled as a site of this MRF which is to be labeled with a discrete significance value to reflect its correctness as image interpretation. In contrast to other approaches using MRFs in computer vision, the neighborhood system models not spatial, but rather supporting or competing relations between grouping hypotheses. While the former is given as the part-of relation, the latter represents contradiction between different interpretations of image data. A prior knowledge about the interaction between groupings is modeled with pair-site clique potentials for both types of neighborhoods. To capture data dependencies in the a posteriori energy we define the data driven significance. It measures the compliance of a given hypothesis to the underlying Gestalt principle in a quantitative way, as opposed to a hard decision as employed in the generation stage. Minimization of the resulting energy integrates supporting and competing interactions between hypotheses as well as the employed Gestalt principles in an uniform framework. Selection of hypotheses on the basis of their final significance completes our grouping process. Extensions of the hierarchy, as mention at the beginning of this section, can be incorporated into both the generation and judgment stage again rather straightforward: Means for hypothesizing groupings of the new types have to be supplied and the data driven significance must be defined. This leaves only definitions for competing relations for the neighborhood system of the grouping graph to achieve integration into the proposed grouping system.

As demonstrated on real images from a construction domain, the resulting significant contour groups represent salient structure of the image and explicitly represent important closed contours as well. Fragmentation due to low contrast and especially occlusion can be resolved in many cases. In addition, integration of cues from region segmentation reduces drawbacks inherent to contour-based approaches. Summarizing, the grouping process with cue integration is successful in exploiting regularities of image data using principles of perceptual organization and thus yields more robust and abstract intermediate segmentation results to facilitate further processes like object recognition or 3D reconstruction.

# References


[1] Arnon Amir and Micheal Lindenbaum. A generic grouping algorithm and its quantitative analysis. *Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 20(2):168–185, 1998.

[2] J. Benois and D. Barba. Image segmentation by region-contour cooperation for image coding. In *Proc. International Conference on Pattern Recognition*, volume C, pages 331–334, 1992.

[3] Julian Besag. Spatial interaction and statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36:192–236, July 1974.

[4] M. Boldt, R. Weiss, and E. Riseman. Token-based extraction of straight lines. *IEEE Trans. on Systems, Man, and Cybernetics (SMC)*, 19:1581–1594, 1989.

[5] J.F. Canny. Finding edges and lines in images. Technical Report 720, MIT Artificial Intelligence Laboratory, 1983.

[6] Paul B. Chou and Christopher M. Brown. The theory and practice of bayesian image labeling. *Int. Journal of Computer Vision*, 4:185–210, 1990.

[7] P.R. Cooper, S. Hyun, and P. Yuen. A markov random field model of subjective contour perception. In *Proc. International Conference on Pattern Recognition*, volume D, pages 100–104, 1996.

[8] C. Distler, D. Boussaoud, R. Desimone, and L.G. Ungerleider. Cortical connections of inferior temporal area teo in macaque monkeys. *Journal of Comparative Neurology*, 334:125–150, 1993.

[9] John Dolan and Edward Riseman. Computing curvilinear structure by token-based grouping. In *Proc. Computer Vision and Pattern Recognition*, pages 264–270, 1992.

[10] H. Edelsbrunner and E. Mücke. Three–dimensional alpha shapes. *ACM Transactions on Graphics*, 13:43–72, 1994.

[11] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 6:721–741, 1984.

[12] Gideon Guy and Gerard Medioni. Perceptual grouping using saliency-enhancing operators. In *Proc. 11. Int. Conf. on Pattern Recognition (ICPR)*, volume I, pages 99–103, Den Haag, 1992.

[13] Olof Henricsson and Markus Stricker. Exploiting photometric and chromatic attributes in a perceptual organization framework. In *Proc. 2nd Asian Conference on Computer Vision*, volume III, pages 258–262, Singapore, 1995. IEEE.

[14] Daniel P. Huttenlocher and Peter C. Wayner. Finding convex edge groupings in an image. In *Proc. Computer Vision and Pattern Recognition*, pages 406–412, Lahaina, Maui, 1991. IEEE.

[15] Alain Jacot-Descombes and Thierry Pun. Asynchronous perceptual grouping: From contours to relevant 2-d structure. *Computer Vision and Image Understanding*, 66(1):1–24, 1997.

[16] C. Jaynes, F. Stolle, and R. Collins. Task driven perceptual organization for extraction of rooftop polygons. In *Image Understanding Workshop*, pages 359–365, 1994.

[17] H.B. Kang and E.L. Walker. Multilevel grouping: Combining bottom-up and top-down reasoning for object recognition. In *Proc. International Conference on Pattern Recognition*, volume A, pages 559–562, 1994.

[18] Peter Kohlhepp. Combining region and edge based segmentation for robust 3d localization in range images. In *Proc. 2nd Asian Conference on Computer Vision*, volume 2, pages 254–258, Singapore, 1995. IEEE.

[19] S. Krishnamachari and R. Chellappa. Delineating buildings by grouping lines with mrf's. *Trans. on Image Processing*, 5:164–168, 1996.

[20] F. Kummert, G.A. Fink, G. Sagerer, and E. Braun. Hybrid Object Recognition in Image Sequences. In *14th International Conference on Pattern Recognition*, volume II, pages 1165–1170, Brisbane, 1998.

[21] A. Leonardis, A. Gupta, and R. Bajcsy. Segmentation of range images as the search for geometric parametric models. *International Journal of Computer Vision*, 14:253–277, 1995.

[22] Aleš Leonardis. *Image Analysis Using Parametric Models.* PhD thesis, University of Ljubljana, 1993.

[23] S.Z. Li. *Markov Random Field Modeling in Computer Vision.* Springer, 1995.

[24] David G. Lowe. *Perceptual Organization and Visual Recognition.* Kluwer Academic Press, Boston, 1985.

[25] A. Maßmann and S. Posch. Mask-oriented grouping operations in a contour-based approach. In *Proc. 2nd Asian Conference on Computer Vision*, volume 3, pages 58–61, Singapore, 1995. IEEE.

[26] James D. McCafferty. *Human and Machine Vision. Computing Perceptual Organisation.* Ellis Horwood, New York, 1990.

[27] Rakesh Mohan. *Perceptual Organization for Computer Vision.* PhD thesis, IRIS Report 254, Institut for Robotics and Intelligent Systems, Univ. of Southern California, 1989.

[28] Rakesh Mohan and Ramakant Nevatia. Perceptual organization for scene segmentation and description. *Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 14(6):616–635, 1992.

[29] V. Murino, C.S. Regazzoni, and G.L. Foresti. Grouping as a searching process for minimum-energy configurations of labeled random-fields. *CVIU*, 64(1):157–174, July 1996.

[30] H. Niemann and R. Salzbrunn. Repräsentation und Nutzung geometrischer Bezüge für die Objekterkennung. In W. Schwerdtmann, editor, *Ausgewählte Verfahren der Mustererkennung und Bildverarbeitung*, pages 52–65. VDI-Verlag, 1989.

[31] E.J. Pauwels, P. Fiddelaers, and L.J. Van Gool. Autonomous grouping of contour-segments using an adaptive region-growing algorithm. In *Proc. International Conference on Pattern Recognition*, volume B, pages 586–590, 1996.

[32] Theo Pavlidis and Yuh-Tay Liow. Integrating region growing and edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-12(3):225–233, March 1990.

[33] Stefan Posch. *Perzeptives Gruppieren und Bildanalyse*. Habilitationsschrift, Universität Bielefeld, 1997.

[34] Lutz Priese and Volker Rehrmann. A fast hybrid color segmentation method. In S. Pöppel and H. Handels, editors, *Mustererkennung 1993. Proceedings 15. DAGM-Symposium*, pages 297–304. Springer, 1993.

[35] G. Rickheit and I. Wachsmuth. Collaborative research centre “Situated Artificial Communicators” at the University of Bielefeld, Germany. *Artificial Intelligence Review*, 10(3-4):165–170, 1996.

[36] Irvin Rock and Stephen Palmer. The legacy of Gestalt psychology. *Scientific American*, 263(6):84–94, December 1990.

[37] Thomas Sanocki, Kevin W. Bowyer, Michael D. Heath, and Sudeep Sarkar. Are edges sufficient for object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1):1–10, 1998.

[38] Sudeep Sarkar and Kim L. Boyer. *Computing Perceptual Organization in Computer Vision*. World Scientific Publishing, 1994.

[39] Amnon Shashua and Shimon Ullman. Grouping contours by iterated pairing network. In *NIPS*, pages 335–341, 1990.

[40] Anne Treisman, Patrick Cavanagh, Burkhart Fischer, V.S. Ramachandran, and Rüdiger von der Heydt. Form perception and attention. In Lothar Spillmann and John S. Werner, editors, *Visual Perception. The Neurophysiological Foundations*, chapter 11. Academic Press, San Diego, 1990.

[41] Shimon Ullman and Amnon Sha'ashua. Structural saliency: The detection of globally salient structures using a locally connected network. Technical report, A.I. Memo 1061, 1988.

[42] Sabine Urago, Josiane Zerubia, and Marc Berthod. A markovian model for contour grouping. In *Proc. International Conference on Pattern Recognition*, volume A, pages 556–558, 1994.

[43] Sabine Urago, Josiane Zerubia, and Marc Berthod. A markovian model for contour grouping. *Pattern Recognition*, 28(5):683–693, 1995.

[44] Max Wertheimer. Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung*, 4:301–350, 1923.

[45] Andrew P. Witkin and Jay M. Tenenbaum. On the role of structure in vision. In Jacob Beck, Barabara Hope, and Azriel Rosenfeld, editors, *Human and Machine Vision*, pages 481–543. Academic Press, Orlando, 1983.

[46] Yu Xiaohan and Juha Ylä-Jääski. Image segmentation combining region growing and edge detection. In *Proc. International Conference on Pattern Recognition*, volume III, pages 481–484, 1992.

[47] J. Xuan, T. Adali, and Y. Wang. Segmentation of magnetic resonance brain image: Integrating region growing and edge detection. In *Proc. International Conference on Image Processing*, volume C, pages 544–547, 1995.

[48] A. Ylä-Jääski and F. Ade. Grouping symmetrical structures for object segmentation and description. *Computer Vision and Image Understanding*, 63:399–417, 1996.