Int. Journal Pattern Recognition and Image Analysis, 1997, 11-17

Perceptual Grouping of Contour Segments Using Markov Random Fields

F. Ackermann, A. Maßmann, S. Posch, G. Sagerer, D. Schlüter

AG für Angewandte Informatik

Universität Bielefeld

Postfach 100131, 33501 Bielefeld

Germany

email: {massmann | posch}@techfak.uni-bielefeld.de

Abstract

The aim of this work is to exploit regular structure in a scene by using the gestalt laws of perception in the field of computer vision. The statistical result of a hand labelled training set is employed to derive "Areas of perceptual attentiveness". Grouping hypotheses are thus generated based on local evidence. To judge these hypotheses in a more global context a Markov random field is used. The approach is contour-based and different types of grouping hypotheses define a hierarchy according to their growing complexity.

1 Introduction

Witkin and Tenenbaum postulate in their paper "On the role of structure in vision" [Tenenbaum and Witkin, 1983] that any appearance of spatiotemporal coherence or regularity is unlikely to arise by chance. This implies an underlying causal relation which for example can be described by an object. They called this the "principle of non-accidentalness". Human beings are able to spontaneously perceive the salient structures even of an unknown scene. This shows that the perception of structure is a relevant mechanism for our daily life. The principles of this so called "perceptual grouping" were described by Gestalt-psychologists already in the twenties (see [Köhler, 1971], [Katz, 1969], [Metzger, 1986], [Rock and Palmer, 1991]). In figure 1 a typical scene of our domain motivates our aim to introduce these structuring mechanisms. A lot of small fragmented contour segments can be recognised due to the wooden texture of the surface. Nevertheless a human being is able to spontaneously perceive the salient structures in the scene: some rectangles, circles, and a hexagon. To exploit these regularities for scene interpretation means to simultaneously reduce ambiguous segmentation data and scene complexity. In [Mohan, 1989] a contour based system for specific collated features (straight lines) and collated features for generic shapes is introduced. Specific features are grouped into lines, parallels, u-contours and rectangles. A constraint satisfaction network accomplishes the global judgement step. Features with a "part-of" relation are supporting like parallels and u-contours. A competing relation exists between equal features. In [Sakar and Boyer, 1991] the grouping process is organized into different levels of a hierarchy. At each level of this hierarchy voting methods, graph operations, and knowledge-based reasoning with a modified Bayesian network formalism take place. The hierarchy levels are derived from the different complexity of the created groupings.

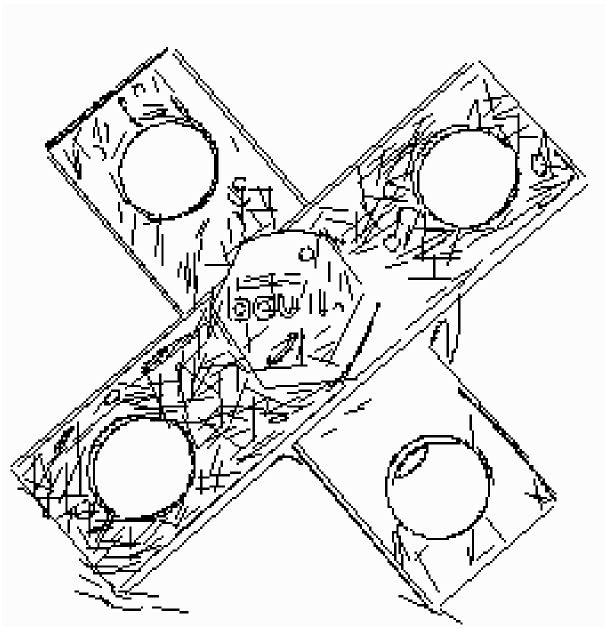

Figure 1: The "air-screw" of a toy plane: a typical scene from our domain, showing salient structures together with fragmented "contour rubbish"

2 The Gestalt laws of perception

Gestalt-psychology is far more than a description of the human visual perception. The research work was done in a broad field like perception, productive thinking, problem solving, memory, and thinking (see also [Sakar and Boyer, 1993]). In computer vision it is appropriate to concentrate upon a summing up of different relations as shown in figure 2. Thus the Gestalt theory can be simplified as the following principles:

- *Proximity* means that things which lie close together are grouped together
- Things which look *similar*, like the black and white rectangles, are grouped together
- A *closed* figure is preferred to an open figure, thus the displayed brackets tend to be perceived as squares.
- Smooth *continuities* are preferred by the visual system, thus the figure is perceived as two intersecting lines (shown left) instead of two cusps (shown right)
- *Symmetrical* areas tend to be perceived as figures
- The larger of the two areas is perceived as *ground* and the white part inside as *figure*
- Things which lie within a *common region* or are *connected* are grouped together

3 A hierarchical contour-based approach

In this work we also 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 organisation. A contour based approach as described here is not able to make use of every gestalt law. Thus we concentrate on proximity, good continuation, symmetry, and closure relations. Good continuation is referred to as collinearity for line segments and curvilinearity for elliptic segments. Parallelism is referred to as a special form of symmetry. Closure assumptions are formed from proximities or symmetry relations.

After edge detection and a thinning step edge elements are approximated with straight lines and elliptic arcs (see [Leonardis, 1993, Taubin, 1991]). These

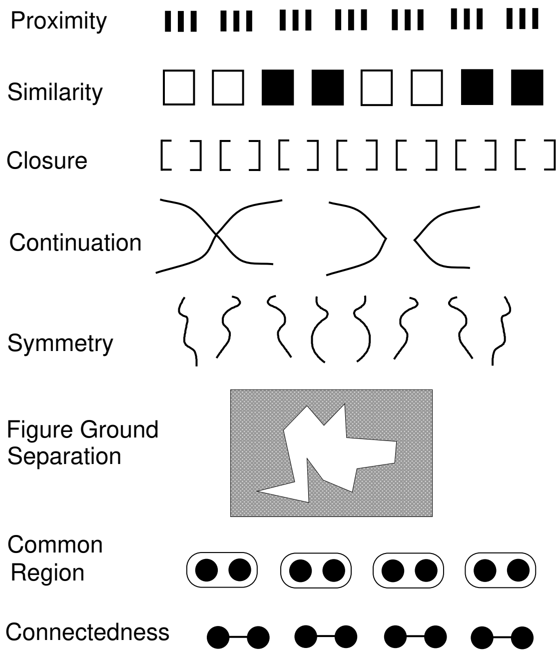

Figure 2: Gestalt laws of perception, the distinct relations are demonstrated due to McCafferty, Rock & Palmer

contour segments are often fragmented due to possible occlusions of different objects or illumination effects. Obvious fragmentations are eliminated using a simple scheme in cases of no ambiguity. The resulting contour segments are the basis for the generation of grouping hypotheses.

Figure 3 shows our hierarchy of different grouping hypotheses. The lowest level consists of collinear, curvilinear, and proximity grouping hypotheses. The medium level describes symmetry and parallelism. The final level consists of closed contour assumptions of two different types: one due to symmetry plus endpoints connected to further segments, which we call a ribbon, and the other without any symmetry or parallelism. The complexity of the different grouping hypotheses is growing within this hierarchy. For example a collinear grouping hypothesis is the basis for a parallelism assumption and a parallel hypothesis is the basis for a closure assumption. Therefore we pursue a hierarchical approach for the grouping process. The naming convention 1D, 2x1D, and 2D reflects this growing complexity. Two different relations between the grouping hypotheses are discerned: support and competition. Support takes place between hypotheses in a "part-of" relation. For example, two line segments which support a collinear

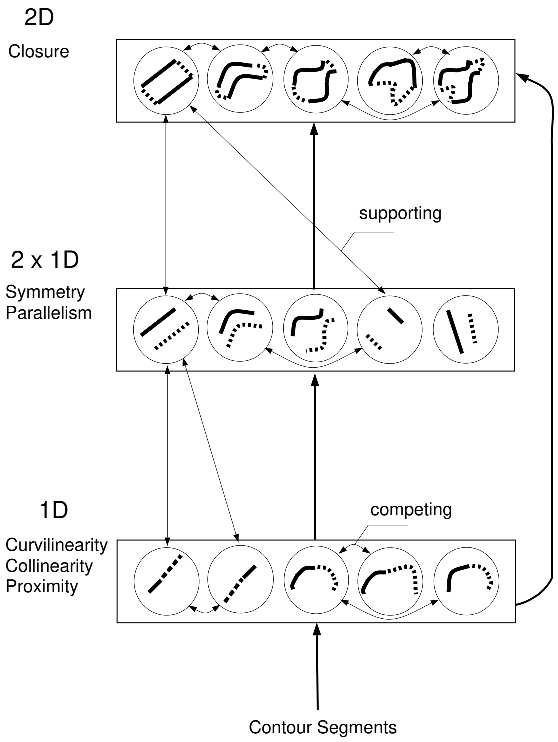

Figure 3: A hierarchical contour-based approach

grouping hypothesis, are each in a part-of relation concerning the collinear hypothesis. No distinct direction is assigned: In our example the line segments support the collinear hypothesis and the collinear hypothesis supports the line segments vice versa. Two different grouping hypotheses which share a common segment constitute an inconsistent interpretation of the image content. These hypotheses are therefore in a competing relation. Competition takes place between hypotheses at the same level of hierarchy . Global judgement is required to solve the ambiguity resulting from competing hypotheses. In figure 3 two parallel hypotheses **support** a closure assumption and two curvilinear hypotheses which share a common segment are **competing**.

4 Areas of perceptual attentiveness

The formation of grouping hypotheses is efficiently done avoiding the combinatorial investigation of all pairs of contour segments. Therefore the search space in the vicinity of endpoints is restricted to a given area (see [Maßmann and Posch, 1995a]), which we call “Area of perceptual attentiveness”. Information about shape and size of these areas is derived for the different grouping operations on the basis of a hand labelled training set of our domain.

(a) (b) (c)

Figure 4: Area of perceptual attentiveness for collinear groupings: a) frequencies of relative endpoint positions, b) positions filtered with Gaussian kernels, c) threshold of filtered "frequency image"

The generation of an area for collinear grouping hypotheses is shown in figure 4. First the endpoint positions of all segments in a collinearity relation are normalised by their length and orientation. This gives the frequency of the endpoint positions. These discrete points are filtered with Gaussian kernels where the variance varies with the frequency of endpoint positions. The result is thresholded to derive an area mask for the interesting collinear search space (see also [Maßmann and Posch, 1995b]).

The generation for example of collinear grouping hypotheses is now recursively accomplished by taking advantage of the area mask. For this purpose the area mask is translated into the distinct endpoints of line segments, rotated according to the orientation of this segment and scaled according to the segment length. The benefit of this operation is an individual search space well matched to the different properties of line segments. A collinear grouping hypothesis for example is created for each segment with an endpoint within the mask. As a further restriction the orientation has to be similar to the continued segment. This collinear hypotheses are recursively expanded by using line segments and collinear hypotheses.

5 Judgement of grouping hypotheses in a global context

With the use of areas of perceptual attentiveness the grouping hypotheses are based on local decisions. A further judgement of these hypotheses in a more global context is necessary to avoid inconsistent groupings. Two main requirements have to be satisfied within this global context:

- The interaction of the locally generated grouping hypotheses which may be competing or supporting
- The detection of the global optimal state by an appropriate control algorithm

A Markov random field is used to find the global optimal state for the locally interacting grouping hypotheses (see [Schlüter, 1995]). The process takes advantage of the neighbourhood system for the formation of local interacting grouping hypotheses. Thus grouping hypotheses are mapped into an undirected graph. Hypotheses correspond to nodes of the graph and edges describe the relations between them. As discussed in section 3 these relations are either competing or supporting. Each node is linked with a random variable. The value of this random variable is a discrete label representing the significance of the given grouping hypotheses. As usual, the energy of a configuration is defined using clique potentials, which in our case model the data dependency and competing or supporting relations between the hypotheses. For each clique of the neighbourhood system a clique potential is defined. Two types of cliques are relevant for our investigations: cliques containing one node (singleton clique) and cliques containing two nodes (doubleton clique). The singleton clique describes the data dependency, the agreement of a grouping hypothesis with its own segmentation data. This could be done by calculating an approximation error. If for example two contour segments are forming a collinear grouping hypothesis a new approximation for the edge pixels of the contour segments as a whole is computed. Besides the approximation error the length of the gap between the two edge chains contributes to the data driven significance. The value for this data driven significance ranges from 0 to 1, where 1 means the maximum agreement with the segmentation data. The actual clique potential for a singleton clique results from the difference between this data-driven significance and the actual significance computed in the neighbourhood system.

The doubleton clique formulates the relations of support and competition. For supporting relations the potential is designed as follows:

- Two significant grouping hypotheses are consistent. In terms of the overall energy minimisation a negative clique potential is assigned.
- A significant hypothesis and a hypothesis with low significance are contradicting. The assigned clique potential is positive.
- Two hypotheses with low significance can neither be interpreted as contradicting or consistent. Therefore the assigned clique potential is zero.

For competing relations the potential is designed vice versa. In figure 5 the discussed relations are displayed in a simplified two-dimensional diagram. The different regions signed with "+" or "-" are approximated with Gaussian curves. The gradients are designed smooth at region boundaries.

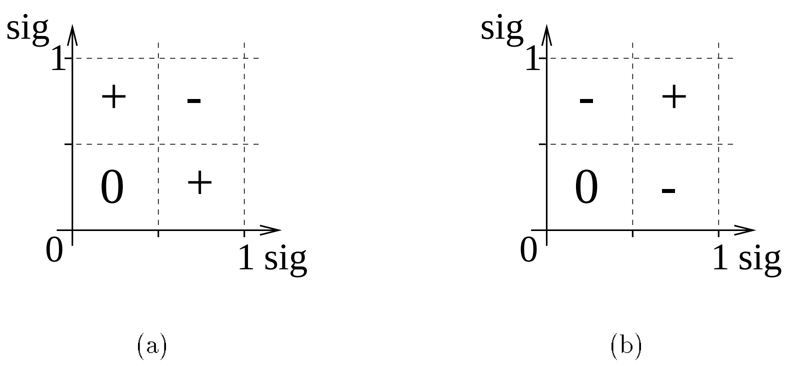

Figure 5: Simplified two-dimensional diagram of the design of the potential functions for doubleton cliques: support(a) and competition(b).

The overall energy minimisation is controlled by "Highest Confidence First" due to Chou & Brown ([Chou et al., 1993]). Chou and Brown describe the advantages of their approach with the key-words: efficiency - predictability - robustness. the iterative procedure is deterministic and performs a maximum improvement at each step. The final estimate depends only upon the inputs and the chosen a priori distribution. The estimates degrade gracefully with the increase of noise and modelling error.

6 Results

On the following pages typical image and segmentation data of our domain (a wooden construction kit) are shown.

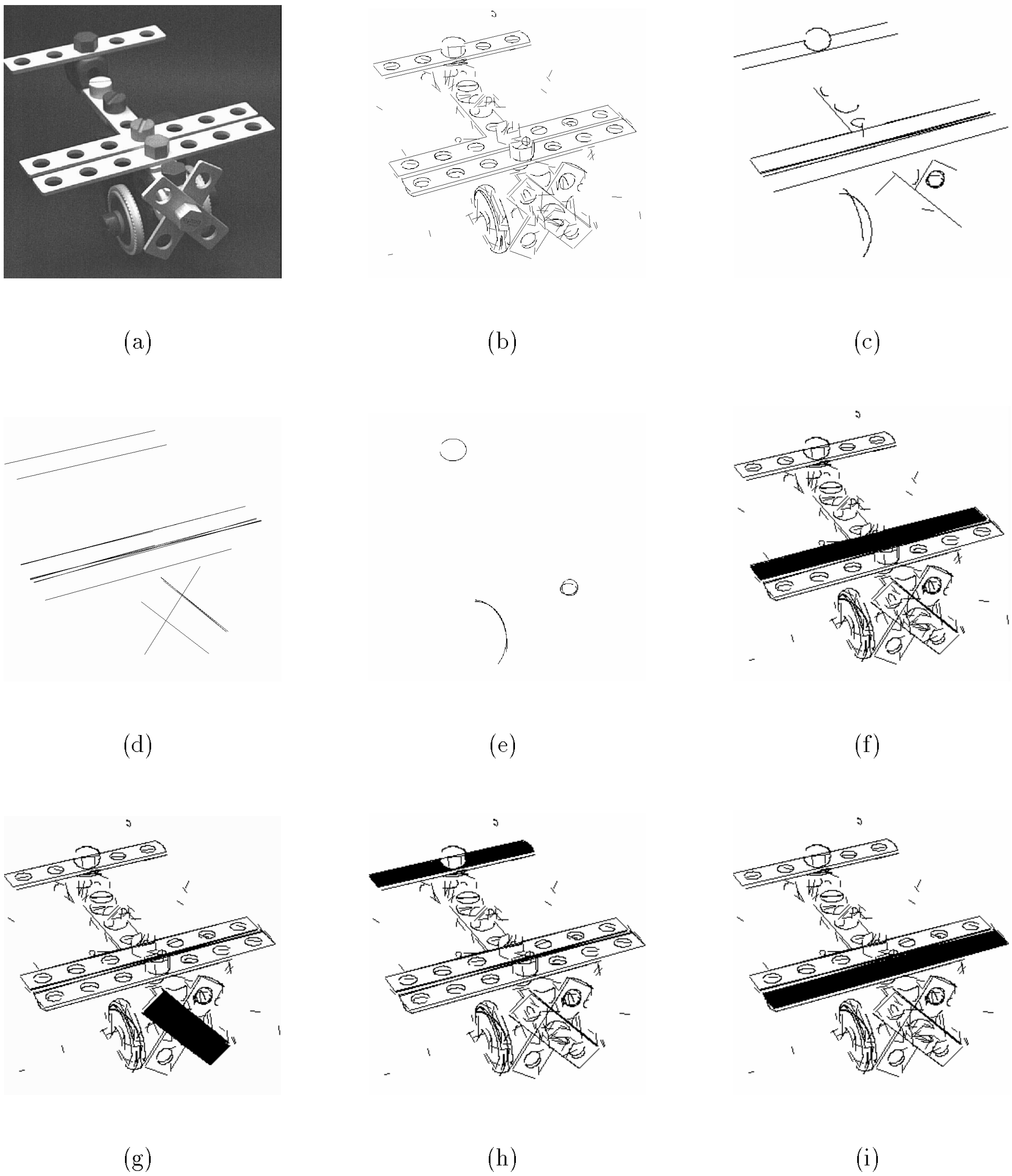

Figure 6: a) wooden toy-plane b) basic contour segments for the grouping step c) all segments and 1D-groupings judged as significant by the Markov random field d) all collinear grouping hypotheses e) all curvilinear grouping hypotheses f,g,h,i) all significant closure hypotheses

The first set of results shows a complete toy plane and the second set shows

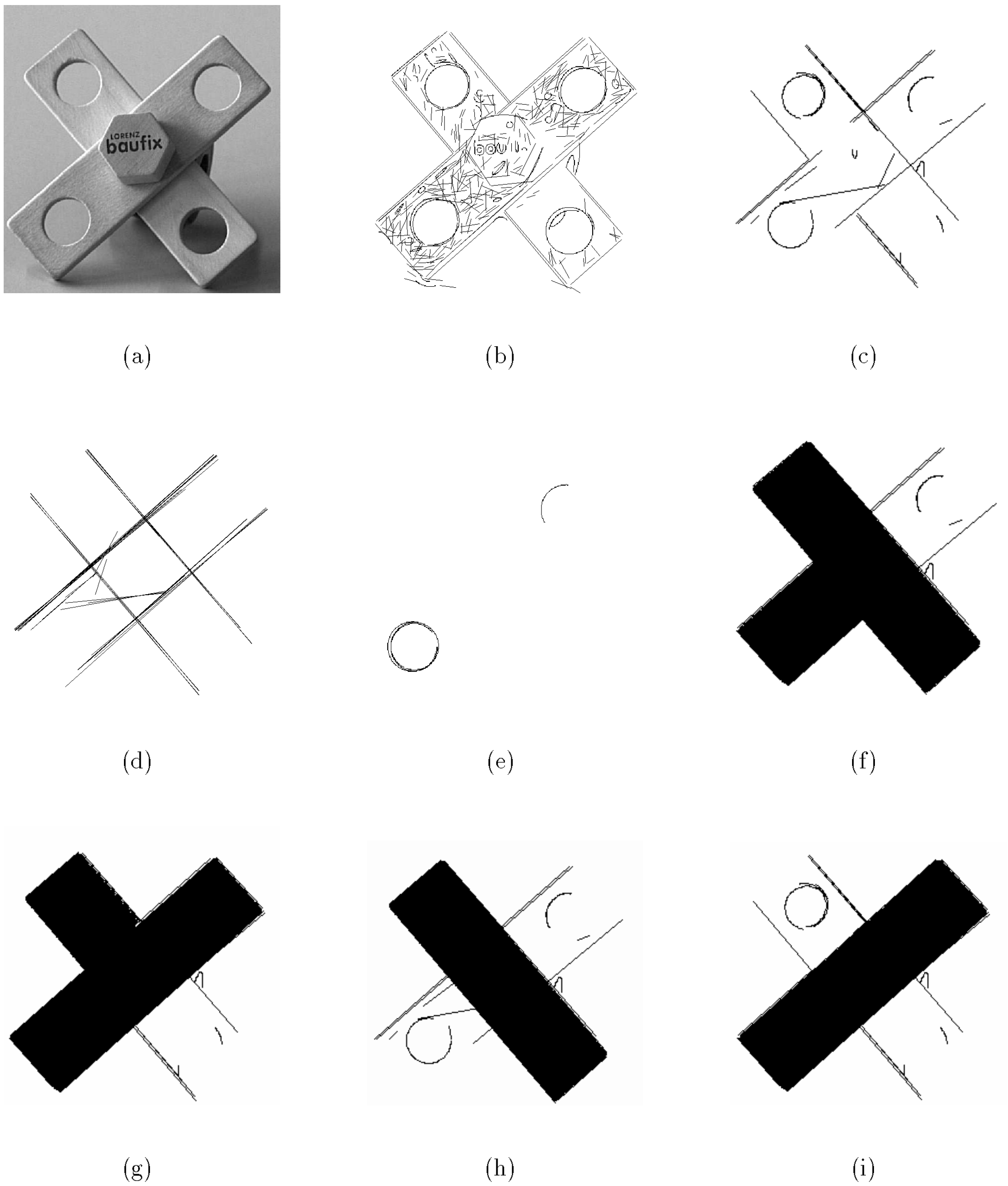

Figure 7: a) air-screw b) basic contour segments for the grouping step c) all segments and 1D-groupings judged as significant by the Markov random field d) all collinear grouping hypotheses e) all curvilinear grouping hypotheses f,g) closure hypotheses with high significance h,i) closure hypotheses with low significance

the air-screw of this plane in more detail (figure 6a, 7a). Both images are of typical scene content for our domain. A toy suggests simple dependencies but nevertheless there are complex structures to cope with. In figure 6b and 7b the segmentation into lines and elliptic arcs is displayed. There are salient structures like parallels, squares, circles, or hexagons besides a great amount of small fragmented contour segments due to illumination effects or texture.

The judgement by the Markov random field results in a significance value for the distinct contour segments and groupings. This value ranges from 0 to 1. A significance value of greater or equal to 0.95 is assumed as significant. The figures 6c, 7c shown these significant segments, collinearities, and curvilinearties. An efficient reduction of the segmentation data is obvious, while most of the salient structures are present. In figures 6d,e and 7d,e all hypothesised collinear and curvilinear groupings are displayed. These hypotheses are formed with the "areas of perceptual attentiveness" (discussed in chapter 4) and fed into the network for further judgement. The remaining four pictures in figures 6, 7 show the hypothesised and already verified closure hypotheses. The assumed closures are marked with a black region. In figures 6f-g, all closures are judged as significant by the network. These hypotheses are corresponding to the salient parts of an aeroplane: wings, tail unit, and air-screw. The lower part of the air-screw is missing but nevertheless this is a good result. In figures 7f-g the first two closure hypotheses (f,g) are judged as significant by the network. The desired two hypotheses (h,i) have a significance near to 0. Further improvement of the generation of grouping hypotheses and parameter tuning inside the network should solve this problem. Future work will be done in this field.

7 Conclusion

The aim is to use gestalt laws in the field of computer vision to make implicit organisation of scene data available for further processing. To exploit this implicit organisation for further interpretation means to simultaneously reduce ambiguous segmentation data and image complexity. Areas of perceptual attentiveness were proposed as a method to reduce the complexity of the grouping process. Information about shape and size of these areas is derived on the basis of a hand labelled training set of our domain. Grouping hypotheses are thus formulated based on local decisions. This local decisions to be verified in a global context. To this end we make us of a Markov random field formalism. The process takes advantage of the neighbourhood system typical for a Markov random field. Grouping hypotheses are mapped into an undirected graph where hypotheses correspond to nodes and edges describe the relations between them. The present results are promising and future work will be done to improve the generation of grouping hypotheses and choice of parameters for the judgement step.

This work has been supported by the German Research Foundation (DFG) in the project SFB 360

References

[Chou et al., 1993] Chou, P., Cooper, P., Swain, M., Brown, C., and Wixson, L. (1993). Markov random fields, theory and application. In Chellappa, R. and Jain, A., editors, *Probabilistic Network Inference for Cooperatve High and Low Level Vision*, pages 211–243. Academic Press.

[Katz, 1969] Katz, D. (1969). *Gestaltpsychologie*. Schwabe.

[Köhler, 1971] Köhler, W. (1971). *Die Aufgabe der Gestaltpsychologie*. de Gruyter.

[Leonardis, 1993] Leonardis, A. (1993). *Image Analysis Using Parametric Models*. PhD thesis, University of Ljubljana.

[Maßmann and Posch, 1995a] Maßmann, A. and Posch, S. (1995a). Bereiche perzeptiver Aufmerksamkeit für konturbasierte Gruppierungen. In *Mustererkennung 95*, pages 602–609.

[Maßmann and Posch, 1995b] Maßmann, A. and Posch, S. (1995b). Mask-Oriented Groupings in a Contour-Based Approach. In *Proceedings Second Asian Conference on Computer Vision 95*, volume III, pages 58–61.

[Metzger, 1986] Metzger, W. (1986). *Gestalt-Psychologie*. Kramer.

[Mohan, 1989] Mohan, R. (1989). *Perceptual Organization for Computer Vision*. PhD thesis, University of Southern California.

[Rock and Palmer, 1991] Rock, I. and Palmer, S. (1991). Das Vermächtnis der Gestaltpsychologie. *Spektrum der Wissenschaft*, pages 68–75.

[Sakar and Boyer, 1991] Sakar, S. and Boyer, K. L. (1991). Integration, Inference, and Management of Spatial Information Using Bayesian Networks: Perceptual Organization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):256–274.

[Sakar and Boyer, 1993] Sakar, S. and Boyer, K. L. (1993). Perceptual organization in computer vision: A review and a proposal for a classificatory structure. *IEEE Transaction on Systems, Man, and Cybernetics*, pages 382–399.

[Schlüter, 1995] Schlüter, D. (1995). Bewertung konturbasierter Gruppierungen mit Hilfe von Markov Random Fields. Master's thesis, Universität Bielefeld.

[Taubin, 1991] Taubin, G. (1991). Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with application to edge and range image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(11):1115–1138.

[Tenenbaum and Witkin, 1983] Tenenbaum, J. M. and Witkin, A. P. (1983). On the role of structure in vision. In Jacob Beck, Barbara Hope, A. R., editor, *Human and Machine Vision*, pages 481–543. Academic Press.