Polymorphic Grouping for Image Segmentation

(to appear at the 5th ICCV, 1995)

Claudia Fuchs, Wolfgang Förstner
Institute for Photogrammetry, Bonn University
Nussallee 15, 53115 Bonn, Germany
e-mail: claudia,wf@ipb.uni-bonn.de

Abstract

The paper describes a new approach to image segmentation. It accepts the inherent deficiencies occuring when extracting low-level features and when dealing with the complexity of real scenes. Image segmentation therefore is understood as deriving a rich symbolic description useful for tasks such as stereo or object recognition in outdoor scenes.

The approach is based on a polymorphic scheme for simultaneously extracting points, lines and segments in a topologically consistent manner, together with their mutual relations derived from the feature adjacency graph (FAG) thereby performing several grouping steps which gradually use more and more specific domain knowledge to achieve an optimal image description.

The heart of the approach is 1.) a detailed analysis of the FAG and 2.) a robust estimation for validating the found geometric hypotheses.

The analysis of the FAG, derived from the exoskeleton of the features, allows to detect inconsistencies of the extracted features with the ideal image model, a cell-complex. The FAG is used for finding hypotheses about incidence relations and geometric hypotheses, such as collinearity or parallelity, also between non-neighbored points and lines.

The M-type robust estimation is used for simultaneously eliminating wrong hypotheses on geometric relationships. It uses a new argument for the weighting function.

Keywords: *image segmentation, grouping, robust estimation.*

1 Introduction

Image segmentation notoriously appears to be the bottleneck for tasks such as object recognition, structural stereo or image interpretation. In spite of the many segmentation techniques available, none of them is flexible enough to be applicable in a large range of domains.

1.1 Related Work and Motivation

Segmentation, understood as *image partitioning* in its general form seems to be not meaningful (Fischler and Bolles 1985) as one generally cannot expect all the relevant boundaries to be observable. Segmentation, understood as deriving a not necessarily complete *symbolic image description* (cf. the 'primal sketch' of Marr 1982) leads to various approaches for detecting the basic features, points, edges, and segments, separately, but due to the incompleteness of the resulting image description motivated the research on low level grouping techniques, especially for edges (cf. Horaud and Veillon 1990, Dolan and Riseman 1992, Huddleston and Ben-Arie 1993). The dichotomy of edge and region based techniques formed the basis for integrating both approaches (cf. German and Geman 1984, Pavlidis and Liow 1990, Chu and Aggarwal 1993), which however, either not explicitly integrate point type features, necessary for consistent analysis (cf. Kovalevsky 1989) or start at the grid structure of the image (cf. Geiger and Yuille 1991) not allowing to represent larger context explicitly, e. g. the straightness of edges in a polyhedral world.

The reason for the difficulties in image segmentation seems to result from the need to integrate data and model driven image analysis in an explicit manner: decisions about the image content should be delayed to later stages of the analysis process where more scene knowledge is available (cf. Price 1984, McCafferty 1990 p. 31). The image model therefore needs to provide several *layers* of increasing complexity, which can absorb the increasing amount of (projected) scene knowledge during information aggregation and allows a smooth transition to higher aggregation procedures. This view strongly supports the grouping paradigm introduced by Lowe 1985 and Biederman 1987 and vividly used since (e. g. Dickinson *et al.* 1992, Sarkar and Boyer 1993).

The necessity to integrate scene knowledge on several levels into the image analysis, integrating data and model driven techniques, therefore requires a broader setup of the first steps within the analysis chain and was the motivation to develop a new approach to image segmentation. It accepts the inherent deficiencies occuring when extracting low-level features and when dealing with the complexity of real images.

1.2 Outline of the Segmentation Procedure

The concept is based on an image model (cf. Fig. 1), starting with an ideal continuous image $\tilde{\mathcal{I}} = (\tilde{\mathcal{P}}, \tilde{\mathcal{L}}, \tilde{\mathcal{S}})$ being a cell-complex with segments $\tilde{\mathcal{S}} = \{\tilde{\mathsf{S}}_k\}$ as 2-cells containing a smooth image function $\tilde{g}(x, y)$, lines $\tilde{\mathcal{L}} = \{\tilde{\mathsf{L}}_j\}$ as 1-cells, and points $\tilde{\mathcal{P}} = \{\tilde{\mathsf{P}}_i\}$ as 0-

Figure 1: Model of the ideal (a.) and the real (b.) continuous image, true points $\tilde{\mathsf{P}}_i$ are expected to lie within the point regions P_i, lines similarly. The borders of the feature regions establish neighborhood relations.

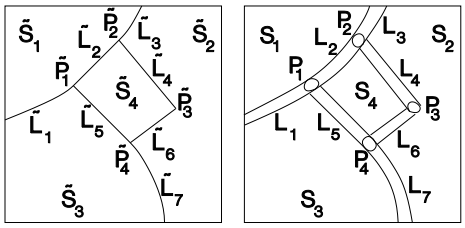
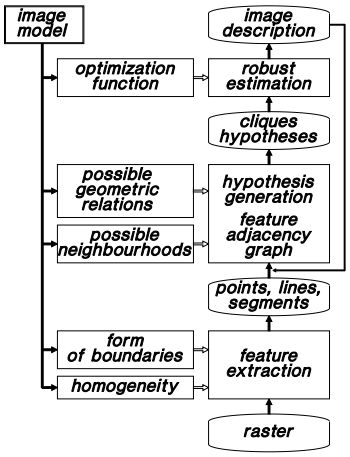


Figure 2: The image segmentation procedure: increasingly more scene knowledge is integrated into the segmentation procedure via the image model. The hypotheses generation and estimation step may be iterated for reaching more complex groupings.



cells, all being open sets (the tilde stands for "true"). The real continuous image $I = (\mathcal{P}, \mathcal{L}, \mathcal{S})$ is a blurred version leading to point and edge regions $\mathcal{L} = \{\mathsf{L}_j\}$ and $\mathcal{P} = \{\mathsf{P}_i\}$ and consequently to segments $\mathcal{S} = \{\mathsf{S}_k\}$ which are smaller in area than the corresponding $\tilde{\mathsf{S}}_k$. The *true discrete image* $g(r, c)$ is a discretized and noisy version of the real continuous image. The task is to derive estimates $\hat{\mathcal{I}}$ and $\hat{g}(x, y)$ of $\tilde{\mathcal{I}}$ and $\tilde{g}(x, y)$ based on the given observations $g(r, c)$ and some generic knowledge about the intensity $\tilde{g}(x, y)$ and the geometry in $\tilde{\mathcal{L}}$ and $\tilde{\mathcal{P}}$. This knowledge in a first instance refers to neighboring features and possibly more complex cliques $\mathcal{C}$ of features which are built by all neighbors of points and segments. The knowledge we use at this stage of the development is primarily of geometric nature, namely straightness, parallelity, rectangularity and collinearity of line features.

Fig. 2 shows the different steps of the segmentation procedure:

1. The *polymorphic feature extraction* (cf. FÖRSTNER 1994) leads to a list of points, lines and segments. It may operate on grey scale or colour images (BRÜGELMANN AND FÖRSTNER 1992) and may lead to straight or curved lines. The low-level image model triggers this step by the chosen homogeneity measure within the segments (shaded grey) and the expected form of the boundaries. An example is given in Fig. (3a, b). Observe that the points and straight line segments are extracted independently and at this stage of the analysis these features represent an accurate but incomplete image description.

2. The *generation of hypotheses* about the geometric properties of the features is based on the feature adjacency graph (FAG) derived from the exoskeleton of the features (cf. Fig. 3c). All neighboring features are tested with respect to a small and sufficient set of geometric relations, taking their uncertainty into account. Focus of attention thereby may be realized by selecting a context dependent set of features.

3. A *M-type robust estimation* fuses the information established so far: the positional information of points and lines and the form information of the hypotheses. The robust estimation simultaneously eliminates possibly wrong or conflicting hypotheses and performs an optimal estimation of the geometry of the segmentation using all information available at that stage (cf. Fig. 3d).

The final result consists of sets of features and geometrical hypotheses which have been proven to be consistent at that level of knowledge. The following steps in the analysis will further increase the percentage of correct hypotheses but can rely on the consistency of the result of the image segmentation.

The setup and the analysis of the FAG for generating hypotheses and the robust estimation are discussed in more detail in the following sections. More examples, also from real images, demonstrate the versatility of the approach.

2 Generating Grouping Hypotheses

The polymorphic feature extraction (FEX) provides a first symbolic description of the image in form of lists of features $\mathsf{F} \in \{\mathsf{P}, \mathsf{L}, \mathsf{S}\}$. The neighborhood analysis of these features based on the feature adjacency graph (Sect. 2.1) allows to identify failures during FEX (Sect. 2.2) and to provide hypotheses for further grouping (Sect. 2.3). Using the FAG for a first interpretation is shown in the examples (Sect. 4).

2.1 The Feature Adjacency Graph

The FAG is an attributed graph $G(\mathcal{F}, \mathcal{N}, \gamma, \rho)$, where the set $\mathcal{F}$ of nodes of G denotes the set of features $\mathcal{F} = \{\mathcal{P}, \mathcal{L}, \mathcal{S}\}$ with geometric and physical attributes γ, and the set $\mathcal{N}$ of edges of G denotes the set of pairs $\mathsf{N}_{ij} \doteq \mathsf{N}(\mathsf{F}_i, \mathsf{F}_j)$ representing neighborhood relations between two features F_i and $\mathsf{F}_j \in \mathcal{F}$ with relational attributes denoted by ρ. Features having a common relation build cliques $\mathcal{C} = \{\mathcal{C}^{(2)}, \mathcal{C}^{(3)}, \ldots\}$ of arbitrary order.

We distinguish two kinds of neighborhood relations: *Direct* neighborhood relations $\mathcal{N}^o = \{\mathsf{N}_{ij}\}$ represent real neighborhoods in the image space. They are defined by the exoskeleton, thus, they are extracted only based on the geometric proximity of the two features in concern. The minimal and the maximal distance and

Figure 3: Roof image (a., pixel size 24 cm at ground level), result of the polymorphic feature extraction (b., points, lines and segments), exoskeleton of the features (c., cf. Fig. 1.b!), and result of segmentation of roof boundaries (d.).

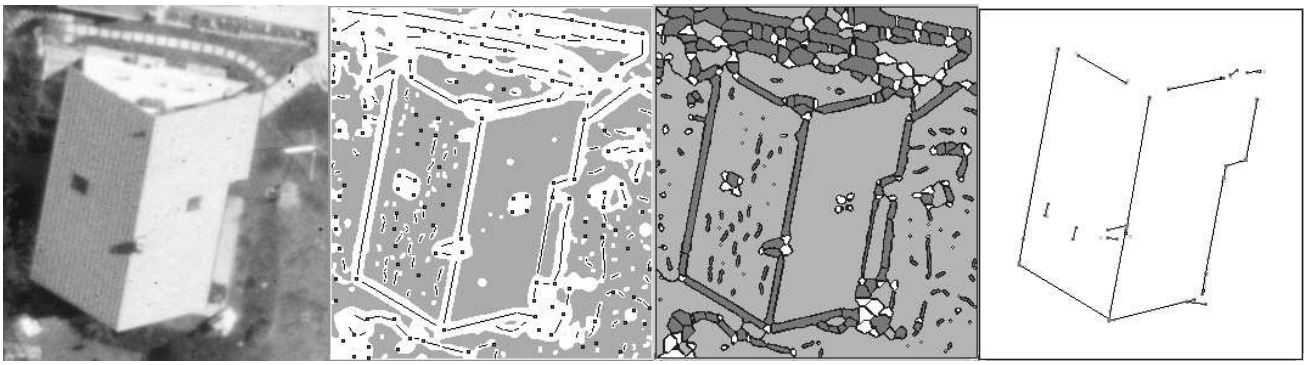
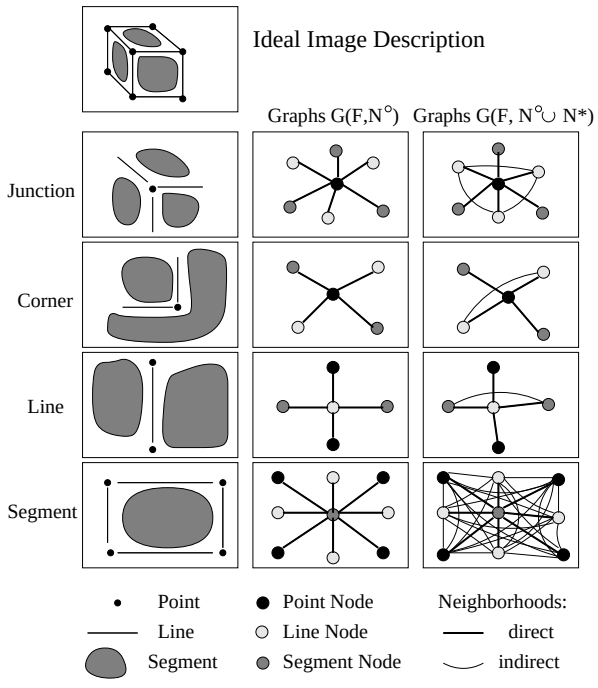


Figure 4: Example of an ideal image description, direct and some useful indirect neighborhood relations.



the length of the common exoskeleton line between two features may be used to characterize their connectivity. Analyzing the exoskeleton has several advantages (EVERS 1988): 1. no threshold is required defining the neighborhood and 2. the complexity of the procedure is independent on the number of features.

In the ideal case (cf. Fig. 4) no neighborhoods of type $\mathsf{N}(\mathsf{P}_i, \mathsf{P}_j)$, $\mathsf{N}(\mathsf{L}_i, \mathsf{L}_j)$, or $\mathsf{N}(\mathsf{S}_i, \mathsf{S}_j)$ occur. As indirect neighbors, especially between segments, separated by a line, and lines separated by a point or common to a segment are important for further analysis, these *indirect* neighborhood relations $\mathcal{N}^*$ are also collected in the FAG, thus $G(\mathcal{F}, \mathcal{N}, \gamma, \rho) = G(\mathcal{F}, \mathcal{N}^0 \cup \mathcal{N}^*, \gamma, \rho)$. Obviously, Fig. 3c reflects the image model Fig. 1b and provides a complete partitioning, however, being polymorphic and the boundaries not carrying geometric information.

The analysis of the FAG can be based on its cliques. We have started the analysis by investigating cliques $\mathsf{C}^{(2)}$ of order 2, thus $\mathsf{C}^{(2)}_{ss}$, $\mathsf{C}^{(2)}_{sl}$, $\mathsf{C}^{(2)}_{sp}$, $\mathsf{C}^{(2)}_{ll}$, $\mathsf{C}^{(2)}_{lp}$ and $\mathsf{C}^{(2)}_{pp}$, in the following omitting the index $^{(2)}$ for brevity.

2.2 Detecting Failures

Fig. 5 shows all types of failures possibly occuring during FEX, which easily can be detected by analyzing the FAG.

Lost and split features cause cliques C_{pp}, C_{ll}, or C_{ss} which do not occur in an ideal segmentation (cf. Fig. 1, 4). The correction, however, is hard, as each of these cliques gives rise to at least two alternatives (cf. Fig. 6 for a situation with C_{ss}).

Spurious features, points or lines, can be hypothesized in case they are isolated within a segment, having only one feature as neighbor, and in case of very short line segments, i. e. less than 4 pixels (cf. FUCHS *et al.* 1994). Obviously, this decision is error prone and does not cover all cases. Therefore, the final decision on spurious features should be postponed, i. e. defining a non-spurious feature to take part in a larger context (cf. below).

Merged features represent an incomplete segmentation and cause inconsistencies with respect to the ideal image model. As can be seen in Fig. 5, again, only a larger context is able to overcome these errors.

Image partitioning does not have difficulties with inconsistencies, which seems to be an advantage. But such schemes hide difficulties, such as the one in Fig. 6, possibly caused by poor contrast, as a decision on one of the both interpretations is made on a too low level, which is difficult to be undone later. Our scheme makes the deficiencies explicit, allowing to use the strong features for either continuing a data driven

Figure 5: Examples of image descriptions, derived by feature extraction on real images leading to inconsistencies with resp. to the image model.

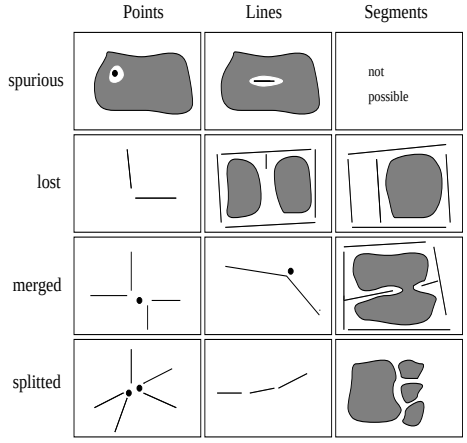
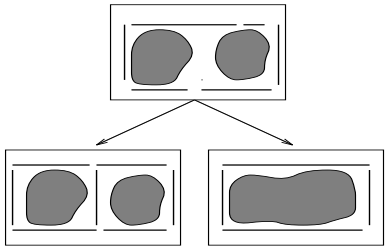


Figure 6: Two possible interpretations of a situation with a neighborhood C_{ss}.



analysis or to initiate higher level hypotheses, in order to gather enough knowledge for resolving the situation.

2.3 Geometric Hypotheses

Interpreting scenes with man made objects allows to introduce general geometric knowledge already at this stage of the image analysis without having to refer to the semantic content of the image. In contrast to more global techniques, like the Hough-transform which requires to fix the 'scale', i. e. the radius of interest, we start with a local analysis using the cliques C_{ll}, C_{lp}, and C_{pp} of the FAG, as points and lines carry the geometric structure of the image.

Table 1 Relations between straight edge segments.

Orientation \ Incidence	endpoint-endpoint	endpoint-line: $\in$ line	endpoint-line: $\notin$ line	no incidence
perpendicular				
parallel				
arbitrary				

Two points P_i and P_j with $\mathsf{N}(\mathsf{P}_i, \mathsf{P}_j)$, a point P_i and the end point of a line L_j with $\mathsf{N}(\mathsf{P}_i, \mathsf{L}_j)$ or two end points of two lines L_i and L_j with $\mathsf{N}(\mathsf{L}_i, \mathsf{L}_j)$ are tested for **identity**. A point P_i and a line L_j with $\mathsf{N}(\mathsf{P}_i, \mathsf{L}_j)$ or a end point of a line L_i and a line L_j with $\mathsf{N}(\mathsf{L}_i, \mathsf{L}_j)$ are tested for **point-line-incidence**. Neighborhood relations between two lines L_i and L_j with $\mathsf{N}(\mathsf{L}_i, \mathsf{L}_j)$ are tested for **collinearity, parallelity** and **orthogonality**. For neighboring lines the 12 semantically different relations collected in Table 1 are hypothesized by technically combining the above mentioned point-point and point-line relations with the angle between the lines. In all cases a hypothesis test on the geometric relation is performed, taking the uncertainty of the features into account (cf. KANATANI 1991).

3 Robust Estimation for Hypotheses Selection and Geometric Reconstruction

3.1 Fusion of Image Data and Hypotheses

The feature extraction and the analysis of the FAG provided us with two types of information, namely the position of the image features and the hypothesized geometric relations. The task now is to find an optimal reconstruction of the underlying true segmentation based on this information. This reconstruction has to take the different type of uncertainty of both information sources into account:

a. Sets $\mathcal{P} = \{\mathsf{P}_i\}$ and $\mathcal{L} = \{\mathsf{L}_j\}$ of points and straight line segments $\mathsf{L}_j = (\mathsf{B}_j, \mathsf{E}_j)$. Their coordinates $\boldsymbol{p} = (x, y)$ are locally estimated. Whereas the coordinates of the points are uncertain, which can be represented by their covariance matrix $\boldsymbol{\Sigma}_{pp}$, we do not assume the features to be wrong, as the percentage of spurious features is quite low (cf. FUCHS *et al.* 1994).

b. A set of geometric hypotheses between these elements. Though the hypotheses are derived using a hypothesis test, they may actually be wrong. Moreover, due to lack of knowledge about the true scene and to approximations in the imaging model (e. g. weak perspective) the hypotheses may even not hold strictly, though they are not false.

A technical solution to the recovery of the true segmentation therefore may be based on a global robust estimation, where only the observed hypotheses are treated as candidates for outliers.

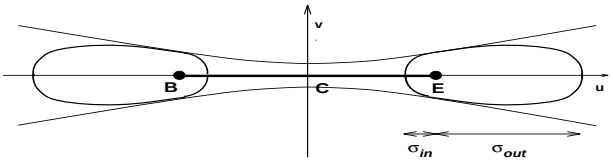
3.2 Model for Geometric Reconstruction

The model for geometric reconstruction may be formalized in the following way:

The functional model relates the expected values $E(\boldsymbol{l})$ of the observed quantities $\boldsymbol{l}$ to the unknown parameters $\boldsymbol{\beta}$ by the in general nonlinear relation $E(\boldsymbol{l}) = \boldsymbol{g}(\tilde{\boldsymbol{\beta}})$ or equivalently the observed values to the estimated parameters $\boldsymbol{l} + \widehat{\boldsymbol{e}} = \boldsymbol{g}(\widehat{\boldsymbol{\beta}})$ with the residuals $\widehat{\boldsymbol{e}} = \boldsymbol{g}(\widehat{\boldsymbol{\beta}}) - \boldsymbol{l}$. The unknown parameters $\widehat{\boldsymbol{\beta}}$ consist of estimates for *all* coordinates of the true feature points or end points of the true line segments.

As the coordinates of the extracted feature points and of the end points of the extracted line segments

Figure 7: Lines of equal probability form 'clubs' not showing full symmetry around begin and end point.



are treated as observations we obtain the very simple functional model $E(x_i) = \tilde{x}_i$ and $E(y_i) = \tilde{y}_i$ or

$$x_i + \widehat{e}_{x_i} = \widehat{x}_i \quad y_i + \widehat{e}_{y_i} = \widehat{y}_i \tag{1}$$

for all coordinates.

As the existence of the relations between the extracted features is uncertain, we introduce these relations as soft constraints, which is done in the following way. All hypotheses may be written as a scalar c or vector $\boldsymbol{c}$ function of the coordinates, whose expected value $E(c)$ or $E(\boldsymbol{c})$ is assumed to be zero. E. g. the *I*ncidence of a point $\tilde{\mathsf{P}}(\tilde{\boldsymbol{p}})$ and a line $\tilde{\mathsf{L}}(\tilde{\mathsf{B}}(\tilde{\boldsymbol{p}}_B), \tilde{\mathsf{E}}(\tilde{\boldsymbol{p}}_E))$ leads to the scalar constraint $c_I(\tilde{\mathsf{P}}, \tilde{\mathsf{B}}, \tilde{\mathsf{E}}) \doteq (\tilde{y}_P - \tilde{y}_B)(\tilde{x}_E - \tilde{x}_B) - (\tilde{x}_P - \tilde{x}_B)(\tilde{y}_E - \tilde{y}_B)$ or the *E*quality of two points to the vector constraint $\boldsymbol{c}(\tilde{\mathsf{P}}(\tilde{\boldsymbol{p}}), \tilde{\mathsf{Q}}(\tilde{\boldsymbol{q}})) \doteq \tilde{\boldsymbol{p}} - \tilde{\boldsymbol{q}}$. In this case it leads to the functional model $\boldsymbol{c}_E(E(\mathsf{P}), E(\mathsf{Q})) = \boldsymbol{c}_E(\tilde{\mathsf{P}}, \tilde{\mathsf{Q}})$ or $\boldsymbol{c}_E(\mathsf{P}, \mathsf{Q}) + \widehat{\boldsymbol{e}}_{c_E} = \boldsymbol{c}_E(\widehat{\mathsf{P}}, \widehat{\mathsf{Q}})$ with $\boldsymbol{c}_E(\mathsf{P}, \mathsf{Q}) = \mathbf{o}$. This explicitly reads as

$$0 + \widehat{\varepsilon}_x = \widehat{x}_Q - \widehat{x}_P \quad 0 + \widehat{\varepsilon}_y = \widehat{y}_Q - \widehat{y}_P \,. \tag{2}$$

Analogously we obtain the observation equations for the other hypotheses.

The stochastical model encodes the uncertainty of the observational values. The reweighting scheme of the robust estimation procedure critically depends on the chosen stochastical properties.

We have to distinguish three types of observations:

1. The stochastical nature of *feature points* is encoded in their 2×2 covariance matrix $\mathsf{D}(\widehat{\boldsymbol{p}}) = \mathsf{D}(x, y) = \boldsymbol{\Sigma}_{\hat{p},\hat{p}}$. It reflects the accuracy of the points with respect to the underlying model, i. e. how well the measured points are expected to fulfill the constraints.

2. The stochastical nature of the *end points* of line segments requires a more detailed modeling. The uncertainty of the end point position *along* the line (in u-direction, cf. Fig. 7) actually is non-symmetric: lines tend to be too short, sometimes quite a bit, and they are seldom too long. The standard deviation $\sigma_{in}(u)$ or $\sigma_{out}(u)$ of u therefore is made dependent on whether the fitted position $\widehat{u}$ is closer or further from the center C of the line segment than u_B or u_E resp. These standard deviations may be derived from the data. In the examples we chose $\sigma_{in} = 1$ pixel and $\sigma_{out} = 10$ pixels restricting to fill short gaps. The standard deviation of the position *across* the edge, thus in v-direction depends on the deviation $\widehat{u}$ from the center C of the line segment. The isolines of the density function on a line segment instead of error ellipses thus are error 'clubs' as in Fig. 7.

3. The stochastical nature of the *constraints* is chosen such a way that they approximate crisp constraints. The variances $\sigma^2_{c_i}$ or covariance matrices $\boldsymbol{\Sigma}_{c_i c_i}$ are related to those, $\overline{\sigma}^2_{c_i}$ and $\overline{\boldsymbol{\Sigma}}_{c_i c_i}$ respectively, derived by error propagation from the observed coordinates. Thus the weights $w_{c_i} = 1/\sigma^2_{c_i}$, $\boldsymbol{W}_{c_i} = \boldsymbol{\Sigma}^{-1}_{c_i c_i}$ of the constraints are k-times larger than the corresponding values derived from the image features. E. g. from $c_{E,x} = x_Q - x_P$ (cf. eq. (2)) follows $\sigma^2_{c_{E,x}} = (\sigma^2_{x_Q} + \sigma^2_{x_P})/k$. We chose $k = 2$. These weights will be modified in the estimation procedure.

The estimation procedure is based on the functional and the stochastical model which can be written as $\boldsymbol{l} + \widehat{\boldsymbol{\varepsilon}} = \boldsymbol{g}(\widehat{\boldsymbol{\beta}})$, $\mathsf{D}(\boldsymbol{l}) = \boldsymbol{\Sigma}_{ll}$. Using the coordinates of the features as approximate values $\boldsymbol{\beta}^{(o)}$ and the Jacobian $\boldsymbol{X} = (x_{ij}) = (\partial g_i/\partial \beta_j)\,|_{\beta_j = \beta_j^{(o)}}$ we obtain the linearized (Gauß Markov) model $\boldsymbol{y} \doteq \boldsymbol{g}(\boldsymbol{\beta}^{(o)}) - \boldsymbol{l} = \boldsymbol{X}\widehat{\Delta\boldsymbol{\beta}}$, $\boldsymbol{W} \doteq \boldsymbol{\Sigma}^{-1}_{yy} = \boldsymbol{\Sigma}^{-1}_{ll}$ or in *groups* of observations

$$\boldsymbol{y}_i \doteq \boldsymbol{g}_i(\boldsymbol{\beta}^{(o)}) - \boldsymbol{l}_i = \boldsymbol{X}_i \cdot \widehat{\Delta\boldsymbol{\beta}}\,,\ \boldsymbol{W}_i = \boldsymbol{\Sigma}^{-1}_{y_i y_i}\,. \tag{3}$$

The estimates $\widehat{\Delta\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}$ lead to better approximations in the next iteration. Generally we have $\widehat{\boldsymbol{\beta}}^{(\nu+1)} = \boldsymbol{\beta}^{(\nu)} + \widehat{\Delta\boldsymbol{\beta}}^{(\nu)}$. The robust estimation consists of an iterative reweighting scheme for the groups of observations

$$\boldsymbol{W}_i^{(\nu)} = \boldsymbol{W}_i^{(0)} \cdot f(A_i^{(\nu)}) \tag{4}$$

using a decreasing weight function $f(\cdot)$ via $A_i^{(\nu)}$. The argument of the weight function $f(\cdot)$ is chosen to be the normalized distance

$$A_i^{(\nu)} = \widehat{\Delta\boldsymbol{y}}_i^{(\nu)} \cdot [\boldsymbol{\Sigma}^{(\nu)}_{\widehat{\Delta y}_i \widehat{\Delta y}_i}]^{-1} \cdot \widehat{\Delta\boldsymbol{y}}_i^{(\nu)}/d^2 \tag{5}$$

The estimated size $\widehat{\Delta\boldsymbol{y}}_i^{(\nu)} = \boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i^{(\nu)[i]} = -\boldsymbol{\Sigma}^{(\nu)}_{y_i y_i} \cdot (\boldsymbol{\Sigma}^{\nu}_{\widehat{e}_i \widehat{e}_i})^{-1}\widehat{\boldsymbol{e}}_i^{(\nu)}$ of the error $\Delta\boldsymbol{y}_i$ in the i-th group of observations is the difference of the observed value $\boldsymbol{y}_i$ and the predicted value $\widehat{\boldsymbol{y}}_i^{(\nu)[i]} = \boldsymbol{X}_i\widehat{\Delta\boldsymbol{\beta}}^{(\nu)[i]}$ in an estimation, where the group $\boldsymbol{y}_i$ of observations is not used (cf. Fig. 8). Its covariance matrix has to be derived by error propagation. The denominator $d^2 = \chi^2_{f,\alpha}$ depends on a significant number and the dimension f of the statistic (1 for scalar, 2 for vector hypotheses).

The choice of argument within a reweighting scheme is new and motivated by the deficiency of classical reweighting schemes. There the residuals are taken as argument, leading to quite unsatisfactory results due to the smearing effects in a least-squares fit. Here the effect of outliers onto the argument of the weight function is completely taken care of.
Convergence usually is achieved after 5-10 iterations.

Figure 8: The argument of the weight function depends on the predicted error $y_i - \widehat{y}_i^{[i]}$ not on $\widehat{e}_i$

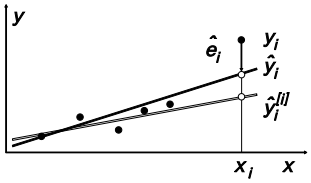
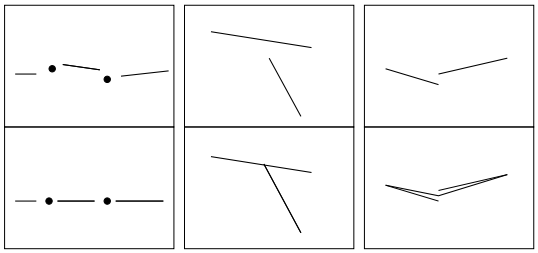


Figure 9: Examples for grouping: top: input, bottom output. Collinearity (a.), point-line incidence (b.) and the effect of the error *clubs* (c.): Instead of the intersection point one obtains a reasonable mean point.



Three examples for the effect of the integration of feature geometry and hypotheses are given in Fig. 9.

The result of the robust estimation consists of a set of constraints which – strictly speaking – could not be disproved by that part of the image model which is used in the estimation procedure. This in our context refers to the set of hypotheses between points and lines, namely incidences and geometric relations. The estimated coordinates of the feature points and the end points of the line segments only are correct if the hypotheses involved actually are correct.

4 Examples

The following examples (cf. Fig. 10) show the result of the proposed polymorphic scheme for deriving a symbolic image description useful for further analysis. All images contain man made objects with mainly polygonal region boundaries. Within the hypotheses generation step short lines (< 6 pixels) and isolated features were ignored in all cases in order to obtain only the strong image structures. The visualization of the result of the robust estimation is primarily meant to show the accepted hypothesis via their effect onto the geometry, and should not be interpreted as the final result of the segmentation.

The first example, taken from ROSENTHALER *et al.* 1992, shows the system to be able to close short gaps and especially the effect of the collinearity constraints along the vertical lines, which are indirectly neighbored via the common feature points. To check for the parallelity constraint is meaningful in this case due to the horizontal viewing direction. The robust estimation analyzed 1103 hypotheses and left 812 (74 %) accepted. As the line features have been assumed to have a positional error (σ) of $0.3\,pels$ across the edge and the hypotheses are introduced as soft constraints the collinearities are not quite fulfilled. Obviously some short, though accepted line segments close to other, longer ones, are prolongated. Partly this appears to be correct, partly not - a situation which cannot be resolved at that level of the analysis.

In the second example, taken from ROTHWELL *et al.* 1993, no hypotheses on parallelity and orthogonality were tested in order to keep the geometry of the weak perspective. 351 (75 %) out of 465 hypotheses were accepted. The positional error of $0.1\,pels$ of the edges lead to nearly no geometric shift or rotation of the edges, however to consistent corner positions. As the current version cannot handle curved boundaries the circular parts of course are erroneously seen as straight. A kind of hysteresis could be applied by first grouping the very strong lines (cf. Fig. 11), here the positional error is $0.3\,pels$) and then link the collinear shorter lines (not shown).

The third example is to show the effect of induced hypotheses. Up to 693 (45 %) out of 1552 hypotheses were accepted, due to the large blobs, causing large cliques with a high percentage of accidental wrong hypothesis. The positional error of the edges was intentionally chosen to be high, $0.5\,pels$. Observe the change of orientation of quite some edges due to parallelity and collinearity constraints, but also the (obviously incorrect) merging of close neighboring lines caused by point-line incidences. Orientations did not change when choosing the error to be $0.3\,pels$.

The last example, taken from the ISPRS test (cf. FRITSCH AND SESTER 1994) is to demonstrate the usefulness of the FAG for goal driven grouping (cf. also Fig. 3). Assuming the purpose of the analysis to be building detection, the selection of the roof boundary features was done based on their color being red, thus using explicit scene knowledge. 66 % (331 out of 505) hypotheses were accepted. The result of the estimation may be used for triggering further analysis, e. g. based on aspects of roof parts.

When evaluating these results one must keep in mind that we did not want to force the local grouping to make decisions which actually require a larger context, be it a scene model or more images.

5 Conclusions

The paper presented an approach for generating a rich symbolic image description. The salient feature of the concept is its openness at the different aggregation levels for changes or extensions in the image model. This refers to the analysis of color or multi-spectral images. Also the low level aggregation to straight line segments may be easily replaced by other schemes, of course requiring the inclusion of, say, circular or conic elements in the robust estimation.

The crucial part is the use of a topologically consistent image model and the attempt to exploit the structure of that model within all analysis steps. This is the cue to detect inconsistencies in the symbolic image description and the guide to build hypotheses for corrections. The proposed feature adjacency graph and its analysis can be transferred to any other feature extraction scheme, as the applied image model, a cell-complex, is rich enough.

Figure 10: Input images (left column), result of FEX (middle column), result of robust estimation (right column). Short lines (< 6 pixels) and isolated features are ignored. Last row: the grouping generation only used features, which have relations to the (red) roof segments

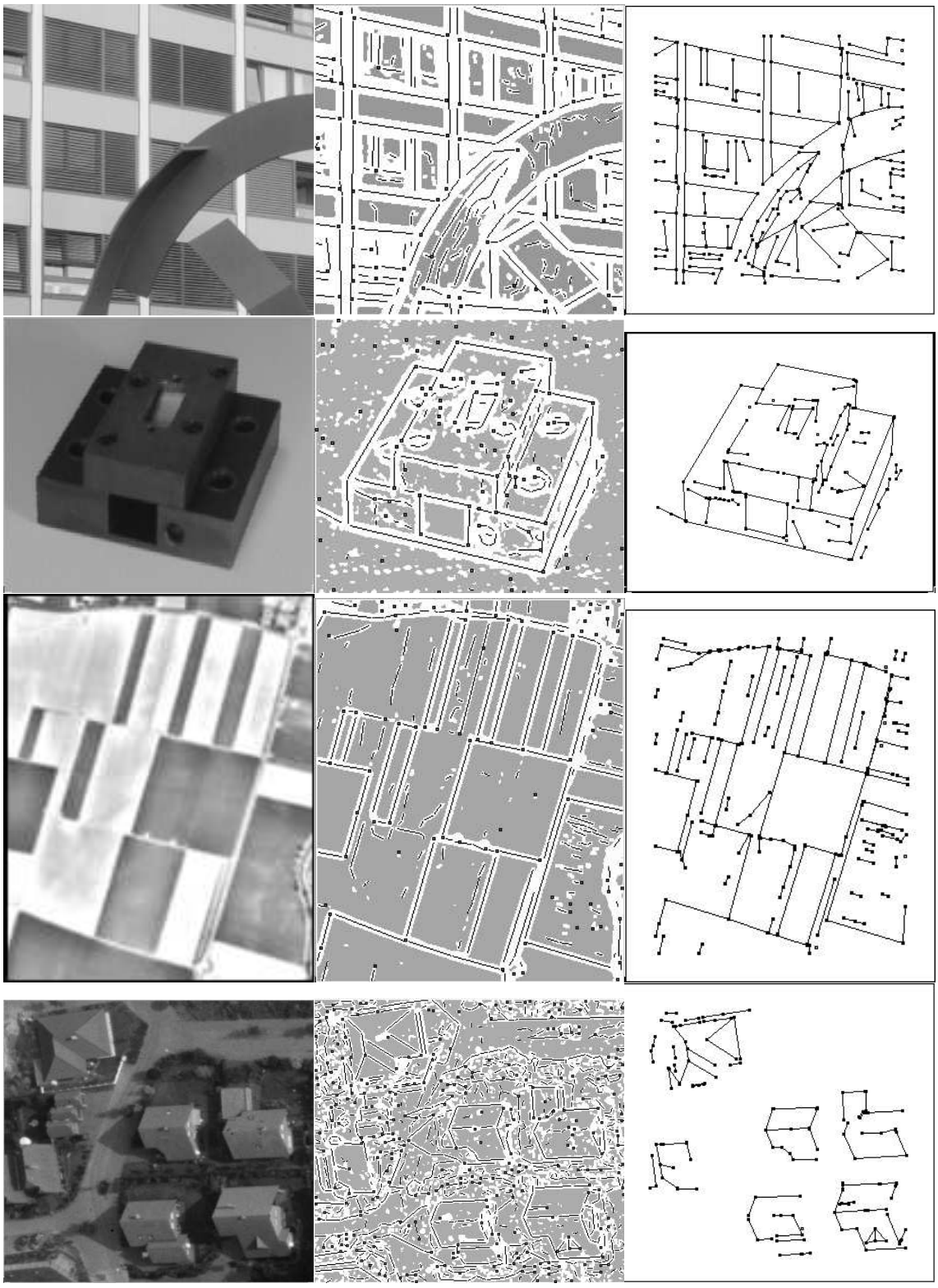
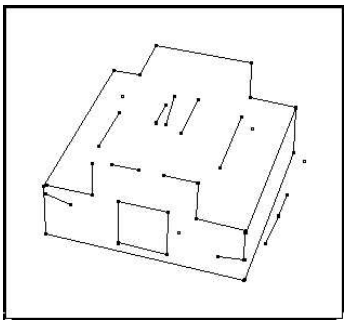


Figure 11: Example 2 from Fig. 10 with only lines $> 15\,pels$.



The experience with the system indicates that the image model, as far as the accuracy of the detected features and the expected relations is concerned, directly influences the result allowing a top-down control of the grouping to be performed. This refers to both, the weighting of the features, representing their expected coherence with the model, and the selection of hypotheses of a certain type. The results on different types of real images show promising results.

The potential of the feature adjacency graph is not yet exploited. The very next step will be to use the local image structures together with the decisions provided by the robust estimation in order to integrate the information of the segments. Furthermore the accepted hypotheses then need to be realized by updating the feature adjacency graph, which then can be used for the next step in the analysis. In case of reconstruction problems, e. g. when using the grouped features in a stereo system using symbolic image matching, the same type of strategy may be applied again, hypothesizing relations and identifying the incorrect ones using a global robust estimation procedure.

References

Biederman, I. (1987): Recognition-by-Components: A Theory of Human Image Understanding. *Psychological Review*, 94(2):115–147, 1987.

Brügelmann, R.; Förstner, W. (1992): Noise Estimation for Color Edge Extraction. In: Förstner, W.; Ruwiedel, S. (Eds.), *Robust Computer Vision*, pages 90–107. Wichmann, Karlsruhe, 1992.

Chu, C.-C.; Aggarwal, J.K. (1993): The Integration of Image Segmentation Maps Using Region and Edge Information. *IEEE T-PAMI*, 15(12):1241–1252, 1993.

Dickinson, S.; Pentland, A.; Rosenfeld, A. (February 1992): 3-D Shape Recovery Using Distributed Aspect Matching. *IEEE T-PAMI*, 14(2):174–198, February 1992.

Dolan, J.; Riseman, E. (1992): Computing Curvilinear Structure by Token-based Grouping. *Proc. of CVPR '92, Champaign, IL*, pages 264–270, 1992.

Evers, Ch. (1988): Analyse von Exoskeletten. *Proc. Mustererkennung 1988, 10. DAGM Symposium München*, 1988. Informatik Fachberichte 180, Springer.

Fischler, M.; Bolles, R. (1985): Perceptual Organization and Curve Partitioning. In: *Image Partitioning and Perceptual Organization*, pages 210–215. IEEE, 1985.

Förstner, W. (1994): A Framework for Low Level Feature Extraction. In: Eklundh, J.-O. (Ed.), *Computer Vision - ECCV 94, Vol. II, Proceedings*, pages 383–394, 1994.

Fritsch, D.; Sester, M. (1994): Test on Image Understanding. In: Ebner H., Heipke C., Eder K. (Ed.), *ISPRS Comm. III Symp.*, pages 243–248. SPIE, 1994.

Fuchs, C.; Lang, F.; Förstner, W. (sep 1994): On the Noise and Scale Behaviour of Relational Descriptions. In: *ISPRS Comm. III Symposium*, pages 257 – 267, Munich, sep 1994.

Geiger, D.; Yuille, A. (1991): A Common Framework for Segmentation. *IJCV*, 6(3):227–243, 1991.

Geman, S.; Geman, D. (1984): Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *Visual System Architectures*, pages 564–584, 1984.

Horaud, R.; Veillon, F. (1990): Finding Geometric and Relational Structures in an Image. In: Faugeras, O. (Ed.), *Computer Vision - ECCV '90*, pages 374–384. Springer-Verlag, 1990.

Huddleston, J.; Ben-Arie, J. (Mar. 1993): Grouping Edgels into Structural Entities Using Circular Symmetry, the Distributed Hough Transform, and Probabilistic Non-accidentalness. *CVGIP: Image Understanding*, 57(2):227–242, Mar. 1993.

Kanatani, K. (Nov. 1991): Hypothesizing and Testing Geometric Properties of Image Data. *CVGIP: Image Understanding*, 54(3):349–357, Nov. 1991.

Kovalevsky, V.A. (1989): Finite Topology as Applied to Image Analysis. *CVGIP*, 46:141–161, 1989.

Lowe, D.G. (1985): *Perceptual Organization and Visual Recognition*. Kluwe, 1985.

Marr, D. (1982): *Vision*. W.H. Freeman, 1982.

McCafferty, J.D. (1990): *Human And Machine Vision*. Ellis Horwood, 1990.

Pavlidis, T.; Liow, Y. (1990): Integrating Region Growing and Edge Detection. *IEEE T-PAMI*, 12(3):225–233, 1990.

Price, A. (1984): Image Segmentation: A Comment on"Studies in Global and Local Histogram-Guided Relaxation Algorithms. *IEEE*, 1984.

Rosenthaler, L.; Heitger, F.; Kübler, O.; von der Heydt, R. (1992): Detection of General Edges and Keypoints. In: *Computer Vision – ECCV '92, Proceedings*, pages 78–86. Springer, 1992.

Rothwell, C.A.; Zissermann, A.; Mundy, J.L. (1993): Extracting Projective Structure from Single Perspective Views of 3D Point Sets. *Proc. of the IV. ICCV, May 11-14,1993, Berlin, Germany*, pages 573–582, 1993.

Sarkar, S.; Boyer, K. (Mar. 1993): Integration, Inference, and Management of Spatial Information Using Bayesian Networks: Perceptual Organization. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 15(3):256–274, Mar. 1993.