

INSTITUT FÜR INFORMATIK  
UND PRAKTISCHE MATHEMATIK

**Geometric/Photometric Consensus and  
Regular Shape Quasi-Invariants for  
Object Localization and Boundary  
Extraction**

Josef Pauli

Bericht Nr. 9805

Mai 1998



CHRISTIAN-ALBRECHTS-UNIVERSITÄT  
KIEL

Institut für Informatik und Praktische Mathematik der  
Christian-Albrechts-Universität zu Kiel  
Olshausenstr. 40  
D – 24098 Kiel

**Geometric/Photometric Consensus and Regular  
Shape Quasi-Invariants for Object Localization  
and Boundary Extraction**

Josef Pauli

Bericht Nr. 9805  
Mai 1998

e-mail: [jpa@informatik.uni-kiel.de](mailto:jpa@informatik.uni-kiel.de)

“Dieser Bericht ist als persönliche Mitteilung aufzufassen.”

## Abstract

Polyhedral descriptions of objects are needed in applications like vision-based robotics, e.g. to carry out grasping and assembling tasks. This work presents a novel methodology for the subtask of localizing a three-dimensional target object in the image and extracting the two-dimensional depiction of the boundary. By eliciting the general principles underlying the process of image formation we exhaustively make use of general, qualitative assumptions, and thus reduce the role of object-specific knowledge for boundary extraction. Geometric/photometric consensus principles are involved in a Hough transformation based approach for extracting line segments. The perceptual organization of line segments into polygons or arrangements of polygons, which originate from the silhouette or the shape of approximate polyhedral objects, is based on shape regularities and quasi-invariants of projective transformation. An affiliated saliency measure combines evaluations of geometric/photometric consensus features with geometric grouping features. An ordered set of most salient polygons or arrangements is the basis for locally applying techniques of object recognition or detailed boundary extraction. The generic approaches are demonstrated for technical objects of electrical strap located in real-world cluttered scenes.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	General context of our work . . . . .	3
1.2	Characterization of our methodology . . . . .	4
1.3	Detailed review of relevant literature . . . . .	5
1.4	Outline of the work . . . . .	8
<b>2</b>	<b>Geometric/photometric consensus principles</b>	<b>9</b>
2.1	Hough transformation for line extraction . . . . .	9
2.2	Orientation-consensus between lines and edges . . . . .	10
2.3	Junction-consensus between convergences and corners . . . . .	15
<b>3</b>	<b>Consensus-based structural level grouping</b>	<b>20</b>
3.1	Configuration of Hough peaks for approximate parallel lines . . . . .	20
3.2	Extraction of regular quadrangles . . . . .	22
3.3	Extraction of regular polygons . . . . .	29
<b>4</b>	<b>Consensus-based assembly level grouping</b>	<b>36</b>
4.1	Focusing image processing on polygonal windows . . . . .	36
4.2	Vanishing point constraint of parallel boundary lines . . . . .	39
4.3	Convergence invariance of meeting boundary lines . . . . .	41
4.4	Boundary extraction for approximate polyhedra . . . . .	42
4.5	Sophisticated geometric reasoning for boundary extraction . . . . .	45
<b>5</b>	<b>Discussion and conclusion</b>	<b>49</b>

# 1 Introduction

The introductory section embeds our methodology of object localization and boundary extraction in the general context of purposive, qualitative vision, then presents a detailed review of relevant literature, and finally gives an outline of the paper.

## 1.1 General context of our work

William of Occam (ca. 1285-1349) was somewhat of a minimalist in medieval philosophy. His motto, which is known as *Occam's Razor*, reads as follows: "*It's vain to do with more what can be done with less*". This economy principle is self-evident in the paradigm of *purposive, qualitative vision* [1].

1. The vision system should gather from the images just the relevant information for solving a specific task.
2. The vision procedures should be generally applicable to a category of similar tasks instead of a single specific task.

Ad 1. In the field of vision-based robotics a category of specific tasks can be the robot grasping of technical objects. In most of these manipulation tasks a polyhedral approximation of the target objects is sufficient (see the survey of robot grasp synthesis algorithms of Shimoga [2]). For example, in order to manipulate an object using a parallel jaw gripper it is sufficient to reconstruct from the image a rectangular solid, though round corners or local protrusion can be a property of the object. It is the corporeal characteristic of the robot gripper, which determines the relevant type of shape approximation to describe the geometric relation between gripper and object. For the qualitative reconstruction of these shapes just a certain spectrum of image analysis tools is useful, which additionally depends on the characteristic of the camera system. For example, if a lens with large focal length is applied, then it is plausible to approximate the geometric aspect of image formation by perspective collineations [3]. As straight 3D object boundary lines are projected into approximate straight image lines, one can use techniques for straight line extraction, e.g. Hough transformation [4]. Altogether, the kind of task determines the *degree of qualitiveness* of the information that must be recovered from the images (see also [5] for interesting contributions to qualitative vision). A restriction to *partial recovery* is inevitable for solving a robot task in limited time with minimum *energy* [6]. In this sense the design of *behavior-based robot systems* drives the design of included vision procedures [7].

Ad 2. Applying Occam's Razor to the design of vision procedures also means to quest for applicability for a *category of similar tasks* instead of a single specific task. This should be reached by exploiting *ground truths* concerning the situatedness and the corporeality of the robot vision system. The ground truths are constraints on space-time, the camera system, and their relationship, which can be generally assumed for the relevant category of robot tasks. General assumptions of various types have been applied more or less successfully in many areas of image processing and computer vision. (a) Profiles of step, ramp, or sigmoid functions are used as *mathematical models* in procedures for edge detection

[8]. (b) For the perceptual organization of edges into structures of higher complexity (e.g. line segments, curve segments, ellipses, polygons) approaches of edge linking are applied, which rely on *Gestaltic phenomena* of proximity, similarity, closure, and continuation [9]. (c) Object recognition is most often based on geometric quantities, which are assumed to be *invariant* under the projective transformations used to model the process of image formation [10]. Usually a stratification into euclidean, similarity, affine, and projective transformations is considered with the property that in this succession each group of transformations is contained in the next. The sets of invariants of these transformation groups are also organized by subset inclusion but in reverse order, e.g. cross ratio is an invariant both under projective and euclidean transformation, however length is only invariant under euclidean transformations. (d) Frequently, in real applications the assumption of invariance is too strong and must be replaced by the assumption of *quasi-invariance*. Its theory and important role for grouping and recognition has been worked out by Binford [11]. For example, Gros et al. [12] use geometric quasi-invariants of pairs of line segments to match and model images. (e) Finally, for the reconstruction of 3D shape and/or motion the ill-posed problem is treated using regularization approaches, which incorporate *smoothness and rigidity assumptions* of the object surface [13]. A critical introspection reveals some problems concerning the applicability of all these constraints. For example, Jain [14] has pointed out that smoothness and rigidity constraints of objects must be applied locally to image regions of depicted object surfaces, but the major problem is to find those meaningful areas. Obviously, in real applications the listed assumptions are too general, and should be more directly related to the type of the actual vision task, i.e. the categories of relevant situations and goals. In knowledge-based systems for image understanding it was proposed to make extensive use of domain-specific knowledge [15]. These systems fit quite well to Marr's theory of vision in the sense of striving for *general vision systems* by explicitly incorporating object-specific assumptions [16]. However, the extensive use of knowledge contradicts to Occam's economy principle, and in many applications the explicit formulation of object models is difficult or even impossible.

## 1.2 Characterization of our methodology

Having the purpose of robot grasping and arranging in mind we present a system for localizing approximate polyhedral objects in the image and extracting their qualitative boundary line configurations. The approach is successful in real-world robotic scenes, which are characterized by clutter, occlusion, shading, etc. A *global-to-local strategy* is favoured, i.e. first to look for a candidate set of objects by taking only the approximate silhouette into account, then to recognize target objects of certain shape classes in the candidate set by applying view based approaches, and finally to extract a detailed boundary.

Our approach of localization is to find *salient polygons*, which represent single faces or silhouettes of objects. The saliency of polygons is based on *geometric/photometric consensus features* and on *geometric regularity features*. The first category of features comprises consensus evaluations between geometric and photometric *line features* and geometric and photometric *junction features*. The Hough transformation is our basic technique for extracting line segments and for organizing them into polygons. Its robustness concerning

parameter estimation is appreciated and the loss of locality is overcome by the geometric/photometric consensus principles. For extracting and characterizing junctions a corner detector is used in combination with a rotating wedge filter. The second category of features, involved in the saliency measure of polygons, comprises *geometric quasi-invariants* under projective transformation. Specifically, they describe regularity aspects of 3D silhouettes respective 2D polygons. Examples for regularities are *parallelism, right-angles, reflection-symmetry, translation-symmetry*.

The ordered places of most salient polygons are visited for *special local treatment*. First, a histogram-based indexing approach can be applied for specific object recognition or recognition of certain shape classes (not treated in this paper). Second, a generic procedure can be applied for detailed boundary extraction of certain shape classes, e.g. parallelepipeds. Our approach is to extract *arrangements of polygons* from the images by incorporating a *vanishing point constraint* and a *convergence invariance constraint*, which both originate from general assumptions of projective transformation of regular 3D shapes. A major contribution of our work is that the basic procedure of line extraction, i.e. Hough transformation, and all subsequent procedures are controlled by constraints, which are inherent in the *three-dimensional nature* of the scene objects and inherent in the *image formation principles* of the camera system.

Our system is organized in several procedures for which the relevant assumptions are clearly stated. The assumptions are related to the situatedness and corporeality of the robot vision system, i.e. quasi-invariants of regular shapes under projective transformation and geometric/photometric consensus of image formation. Furthermore, these assumptions are stratified according to *decreasing generality*, which imposes a certain degree of speciality on the procedures. Concerning the objects in the scene our most general assumption is that the object shape is an approximate polyhedron, and an example for a specific assumption is that an approximate parallelepiped is located in a certain area. We follow the claims of Occam's minimalistic philosophy and elicit the general principles underlying the perspective projection of polyhedra, and then implement procedures as general applicable as possible. Based on this characterization of our methodology, relevant contributions in the literature will be reviewed.

### 1.3 Detailed review of relevant literature

Object detection can be considered as a cyclic two-step procedure of localization and recognition [17], which is usually organized in several levels of data abstraction. Localization is the task of looking for image positions where objects of a certain class are located. In the recognition step one of these locations is considered to identify the specific object. Related to the problem of boundary extraction the task of localization is strongly correlated with perceptual organization, e.g. to organize those gray value edges, which belong to the boundary of a certain object. Sarkar and Boyer [18] have reviewed the relevant work in perceptual organization (up to year 1992) and proposed a four-level classification of the approaches, i.e. *signal level, primitive level, structural level, assembly level*. For example, at the signal level pixels are organized into edge chains, at the primitive level the edge chains are approximated as polylines (i.e. sequences of line segments), at the structural level the polylines are combined to polygons, and at the assembly level several

polygons are organized into arrangements. For future research they suggested: *"There is a need for research into frameworks for integration of various Gestaltic cues including non-geometric ones ..."*. In line with that, the same authors present in [19] a hierarchical system for the extraction of curvilinear or rectilinear structures. Regularities in the distribution of edges are detected using voting methods for Gestaltic phenomena of proximity, similarity, smooth continuity and closure. The approach is generic in the sense that various forms of tokens can be treated and represented as graphs, and various types of structures can be extracted by applying standardized graph analysis algorithms.

Our approach more extensively incorporates non-geometric cues, i.e. photometric features, treats closed line configurations of higher complexity including higher level Gestaltic phenomena, and from that defines a saliency measure for different candidates of line organizations.

Frequently, Hough transformation has been used as basic procedure for grouping at the primitive or structural level [4]. Specific arrangements of gray value edges are voting for certain analytic shapes, e.g. straight lines or ellipses. For example, each line of edges creates a peak of votes in the space of line parameters and the task is to localize the peaks. To make the Hough transformation more sensitive one can go back to the signal level and take the orientations of the gray value edges into account. As orientation is a parameter in a polar representation of a line the number of possible line orientations, for which a pixel may vote, can be reduced to the relevant one. The size of this voting kernel influences the sharpness of the Hough peaks [20], i.e. the accuracy of line parameters. The Hough image can be used for grouping at the structural level or even at the assembly level. The problem is to find especially those peaks, which arise from lines belonging to a specific object. In the work of Princen et al. [21] a hierarchical procedure is used, which extracts an exhaustive set of peaks and afterwards selects the relevant subset by applying Gestaltic grouping criteria. Wahl and Biland [22] extract objects from a polyhedral scene by representing an object boundary as a distributed pattern of peaks in the parameter space of lines. Unfortunately, the approach is demonstrated only for synthetic images. Alternatively, Ballard [23] introduced the generalized Hough transformation for the extraction of complex natural 2D shapes, in which a shape is represented in tabular form instead of an analytic formula.

This short review of extensions of the standard Hough transformation gives the impression, that our Hough voting procedure can serve as basis for perceptual organization at all perception levels and for integration of cues from all levels. The greatest weakness of the standard Hough transformation is the loss of locality, e.g. a line can gain support from pixels anywhere along its length from image border to border. Therefore, two or more line segments may be misinterpreted as one line, or short line segments may be overlooked. Consequently, Yang et al. [24] introduce a weighted Hough transformation, in which the connectivity of a line is measured in order to also detect short line segments. Similarly, Foresti et al. [25] extend the Hough transformation to labeled edges. Each label corresponds to a line segment, which is extracted by a classical line following procedure taking connectivity and straightness in the course of edges into account.

Our approach to overcome this problem is to apply two principles of geometric/photometric consensus. The first one is the line/edge orientation-consensus mentioned above, and the second one considers the characterization of gray value corners and therefore is two-



dimensional in nature. Furthermore, locality of boundary extraction is reached by applying a *windowed Hough transformation* within the areas of most salient polygons.

Cho and Meer [26] propose an approach for detecting image regions by evaluating a consensus among a set of slightly different segmentations. Local homogeneity is based on cooccurrence probabilities derived from the ensemble of initial segmentations, i.e. probabilities that two neighbored pixels belong to the same image region. Region adjacency graphs at several levels are constructed and exploited for this purpose. In our work a consensus methodology is introduced for bridging the gap between geometric and photometric features, and for driving image segmentation based on geometric regularity features (of 3D objects), which are quasi-invariants under projective transformation.

Amir and Lindenbaum [27] present a grouping methodology for both signal and primitive level. A graph is constructed, whose nodes represent primitive tokens such as edges, and whose arcs represent grouping evidence based on collinearity or general smoothness criteria. Grouping is done by finding the best graph partition using a maximum likelihood approach. A measure for the quality of detected edge organizations is defined, which could be used as decision function for selectively postprocessing certain groups. In our approach graph analysis is avoided at the primitive level of grouping edges (because it seems to be time-consuming), but is used for detecting polygons at the structural level.

Castano and Hutchinson [28] present a probabilistic approach to perceptual grouping at the primitive and structural level. A probability distribution over a space of possible image feature groupings is determined and the most likely groupings are selected for further treatment. The probabilities are based on how well a set of image features fits to a particular geometric structure and on the expected noise in image data. The approach is demonstrated for two types of low-level geometric structures, i.e. straight lines and bilateral symmetries. Complex symmetrical structures consisting of groups of line segments are extracted in the work of Ylä-Jääski and Ade [29]. In a two-step procedure first pairs of line segments are detected, which are the basic symmetry primitives, and second several of them are selectively grouped along the symmetry axes of the segment pairs. Our approach is a more general in the sense that we treat further types of regularities and symmetries.

In the work of Zisserman et al. [30] grouping is done at all four levels. Line structures belonging to an object are extracted by using techniques of edge detection, contour following and polygonal approximation (signal level, primitive level, structural level). The representation is given by certain invariants to overcome difficulties in recognizing objects under varying viewpoints. These geometric invariants are used to define an indexing function for selecting certain models of object shapes, e.g. certain types of polyhedra or surfaces of revolution. Based on a minimal set of invariant features a certain object model is deduced, and based on that a class-based grouping procedure is applied for detailed boundary extraction (assembly level). For example, under affine imaging conditions the parallelism of 3D lines of a polyhedra also holds between the projected lines in the image. Accordingly, certain lines of the outer border of an object appear with the same orientation in the interior of the silhouette. This gives an evidence of grouping lines for describing a polyhedron. Our approach contributes to this work, in that we introduce some assembly level grouping criteria for boundary extraction of approximate polyhedra. These criteria are the parallelism quasi-invariant, the convergence invariant, and a

vanishing point constraint.

## 1.4 Outline of the work

Section 2 recalls the definitions of standard Hough transformation and orientation-selective Hough transformation for line extraction. Furthermore, the geometric/photometric consensus features are introduced, which are based on the principle of orientation-consensus between lines and edges, and on the principle of junction-consensus between convergences and corners. Section 3 defines regularity features of polygons, which are projective quasi-invariants, i.e. parallelism or right-angle between line segments, reflection-symmetry or translation-symmetry between polylines. These quasi-invariants are combined with the geometric/photometric consensus features in a generic procedure for extracting salient quadrangles or polygons. Section 4 introduces grouping criteria at the assembly level, i.e. the vanishing point constraint and the convergence invariance. These assembly level criteria are integrated with the consensus features and quasi-invariants at the signal, primitive and structural level. Two generic procedures are presented for extracting the arrangements of polygons for approximate polyhedra. Section 5 discusses the results of the work on the basis of 14 assumptions, stated in the previous sections, which are ground truths inherent in the three-dimensional nature of the scene objects and inherent in the image formation principles of the camera system. The usefulness of the methodology is demonstrated for real-world scenes of electrical scrap.

## 2 Geometric/photometric consensus principles

Obviously, the general assumption behind all approaches of boundary extraction is that three-dimensional surface discontinuities must have corresponding gray value edges in the image. Nearly all problems can be traced back to a gap between the geometric and the photometric type of scene representation. This section introduces measures for evaluating a consensus between geometric and photometric features. The one-dimensional feature of straight lines and the two-dimensional feature of a convergence of two straight lines are considered. Exemplary, the Hough transformation is used as basic procedure for line extraction and the SUSAN operator for corner detection.

### 2.1 Hough transformation for line extraction

For representing straight image lines we prefer the polar form (see Fig. 1), which avoids singularities. Let  $P$  be the set of discrete coordinate tuples  $p := (x_1, x_2)$  for the image pixels of a gray value image with  $I_w$  columns and  $I_h$  rows. A threshold parameter  $\delta_1$  specifies the permissible deviation from linearity for a sequence of image pixels.

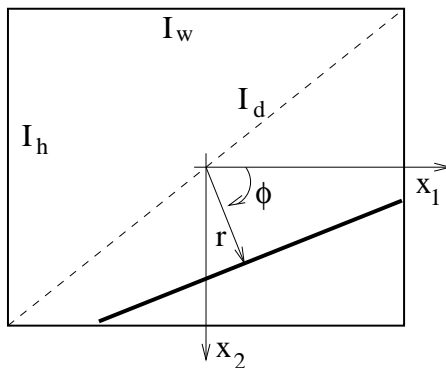


FIG. 1: Polar form of a line with distance parameter  $r$  and angle parameter  $\phi$  taken relative to the image center.

**Definition 1 (Polar representation of a line)** *The polar representation of an image line is defined by*

$$L(p, q) := x_1 * \cos(\phi) + x_2 * \sin(\phi) - r, \quad |L(p, q)| \leq \delta_1 \quad (1)$$

Parameter  $r$  is the distance from the image center to the line along a direction normal to the line. Parameter  $\phi$  is the angle of this normal direction related to the  $x_1$  axis. The two line parameters  $q := (r, \phi)$  are assumed to be discretized. This calls for the inequality symbol in Eq. (1) to describe the permissible deviation from the ideal value zero. For the parameter space  $Q$  we define a discrete two-dimensional coordinate system (see Fig. 2). The horizontal axis is for parameter  $r$ , whose values reach from  $\frac{-I_d}{2}$  to  $\frac{+I_d}{2}$ , where  $I_d$  is the length of the image diagonal. The vertical axis is for parameter  $\phi$ , whose values reach from  $0^\circ$  to  $180^\circ$  angle degrees. The discrete coordinate system can be regarded as a matrix consisting of  $I_d$  columns and 180 rows.

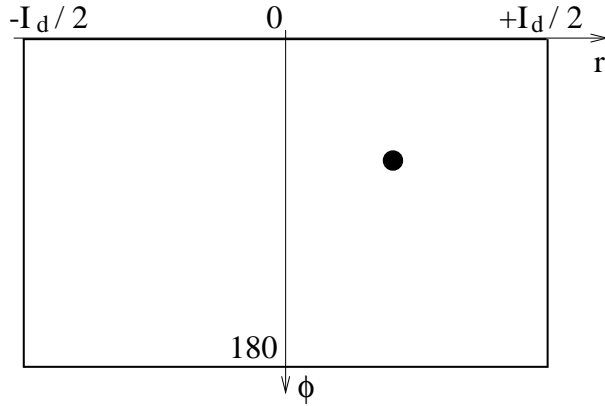


FIG. 2: Coordinate system for the two-dimensional space of line parameters (Hough image); horizontal axis for parameter  $r$  reaching from  $-\frac{I_d}{2}$  to  $+\frac{I_d}{2}$ , and vertical axis for parameter  $\phi$  reaching from 0 to 180.

Due to discretization each parameter tuple is regarded as a bin of real-valued parameter combinations, i.e. it represents a set of image lines with similar orientation and position. The standard Hough transformation counts for each bin how many edges in a gray value image lie along the lines, which are specified by the bin. In the Hough image these numbers of edges are represented for each parameter tuple. Each peak in the Hough image indicates that the gray value image contains an approximate straight line of edges, whose parameters are specified by the position of the peak. A binary image  $B$  is used, which represents the edge points by  $B(p) = 1$  and all the other points by  $B(p) = 0$ .

**Definition 2 (Standard Hough transformation)** *The standard Hough transformation  $SHT : Q \rightarrow [0, \dots, I_w * I_h]$  of the binary image  $B$  relative to the polar form  $L$  of a straight line is defined by*

$$SHT(q) := \#\{p \in P \mid B(p) = 1 \wedge |L(p, q)| \leq \delta_1\} \quad (2)$$

Fig. 3 shows on top the gray value image of a dark box (used as dummy box in electrical equipment) and at bottom the binarized image of gray value edges. The standard Hough transformation is depicted in Fig. 4. Typically, wide-spread maxima occur due to the reason that all edges near or on a line cause the SHT to not only increase the level of the relevant bin but also many in their neighborhood. We are interested in sharp peaks in order to easily locate them in the Hough image (i.e. extract the relevant lines from the gray value image) and accurately estimate the line parameters. By making the discretization of the parameter space more fine-grained the maxima would be more sharpened and more accurate, however the computational expenditure for the Hough transformation would increase significantly.

## 2.2 Orientation-consensus between lines and edges

The conflict between accuracy of extracted lines and efficiency of line extraction can be reduced by making use of an orientation-consensus between lines and gray value edges.

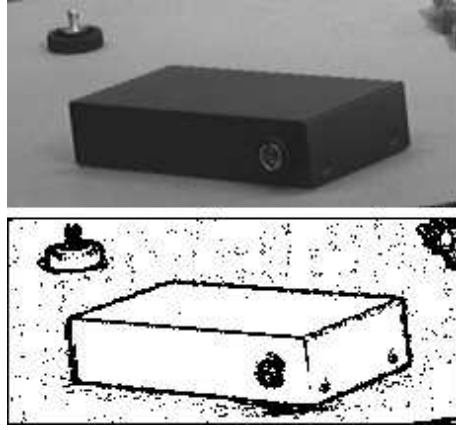


FIG. 3: (Top) Gray value image of an electrical dummy box, (bottom) binarized gradient magnitudes indicating the positions of gray value edges.

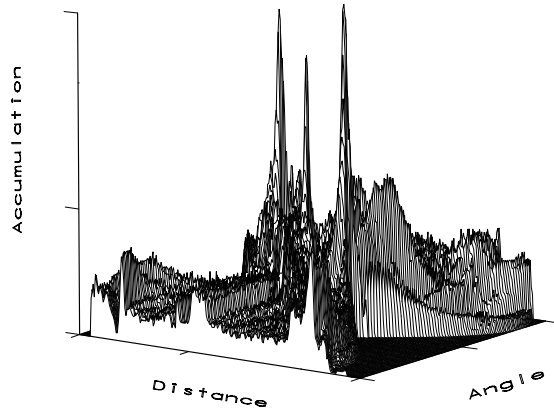


FIG. 4: Standard Hough transformation of the binarized image in Fig. 3. Wide-spread maxima in the Hough image, i.e. the accumulation array with discrete line parameters (distance  $r$  and angle  $\phi$ ) for indexing columns and rows.

**Assumption 1 (Line/edge orientation-consensus for a line point)** *The orientation  $\phi$  of a line of gray value edge points and orientation  $O(p)$  of an edge at point  $p := (x_1, x_2)$  on the line are approximately equal. The replacement of  $\phi$  by  $O(p)$  in the polar form of the image line implies just a small deviation from ideal value zero. The necessary geometric/photometric consensus is specified by parameter  $\delta_2$ .*

$$|x_1 * \cos(O(x_1, x_2)) + x_2 * \sin(O(x_1, x_2)) - r| \leq \delta_2 \quad (3)$$

A line of edge points may originate from the gray value contrast at the object surface, e.g. due to texture, inscription, shape discontinuities or figure/background separation. Small distortions in the imaging process and inaccuracies in determining the edge orientation are

considered in Eq. (3) by parameter  $\delta_2$ , which specifies the upper bound for the permissible error. In our system the orientations of gray value edges are extracted by appropriately combining four differently oriented 2D-Gabor-functions, which are sensitive to the directions  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$  (see [31], pp. 219-258). The standard Hough transformation will be modified by taking the orientation at each edge point into account and accumulating only those small set of parameter tuples, for which Eq. (3) holds. A tolerance band  $\Delta a$  will be introduced to take care for the inaccuracy of the edge orientation  $O(p)$  at position  $p$  (adopted from Princen et al. [21]).

**Definition 3 (Orientation-selective Hough transformation)** *The orientation-selective Hough transformation  $OHT : Q \rightarrow [0, \dots, I_w * I_h]$  of the binary image  $B$  and the orientation image  $O$  relative to the polar form  $L$  of a straight line is defined by*

$$OHT(q) := \#\{p \in P \mid B(p) = 1 \wedge |L(p, q)| \leq \delta_1 \wedge \phi - \frac{\Delta a}{2} \leq O(p) \leq \phi + \frac{\Delta a}{2}\} \quad (4)$$

Fig. 5 shows the result of the OHT, if we assign  $\Delta a = 4^\circ$  angle degrees for the tolerance band of edge orientations. Compared to the SHT in Fig. 4 we realize that more local maxima are sharpened in the OHT.

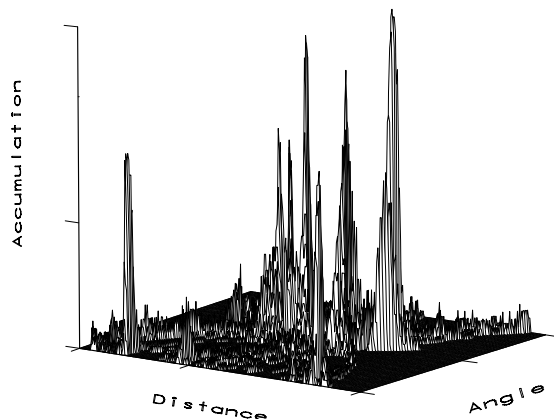


FIG. 5: Orientation-selective Hough transformation of the binarized image in Fig. 3 taking an image of edge orientations into account. Several local maxima are much more sharpened than in the Hough image of SHT in Fig. 4.

The local maxima can be obtained iteratively by looking for the global maximum, erasing the peak position together with a small surrounding area and restarting the search for the next maximum. Due to the sharpness of the peaks in OHT it is much easier (compared to SHT) to control the area size to be erased in each iteration. Fig. 6 shows the extracted lines specified by the set of 10 most maximal peaks in the Hough images of SHT and OHT (top and bottom respectively). Obviously, the lines extracted via OHT, which consider

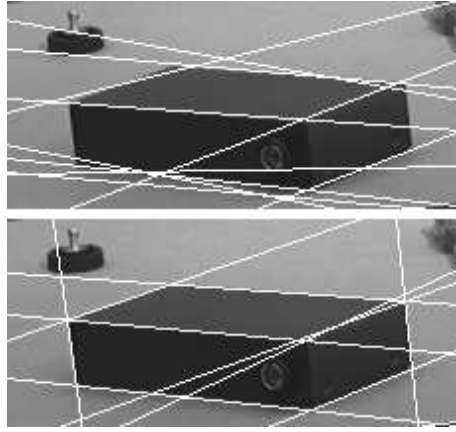


FIG. 6: Extracted image lines based on 10 most maximal peaks in the Hough image, (top) of SHT, (bottom) of OHT. The lines extracted via OHT are more relevant and accurate for describing the object boundary.

the line/edge orientation-consensus, are more relevant and accurate for describing the object boundary.

The line/edge orientation-consensus not only supports the extraction of relevant lines, but is also useful for verifying or adjusting a subset of candidate lines, which are determined in the context of other approaches (see Sections 3 and 4 later on). Furthermore, the orientation-consensus is used to verify certain segments of lines, i.e. restrict the unbounded lines extracted with OHT to the relevant segments of an object boundary. The finite set of discrete points  $p_i, i \in \{1, \dots, N\}$ , of a line bounded between  $p_1$  and  $p_N$  is denoted by  $\mathcal{L}(p_1, p_N)$ . In Fig. 7 the line  $\mathcal{L}(p_a, p_d)$  through the characteristic points  $\{p_a, p_b, p_c, p_d\}$  is only relevant between  $p_b$  and  $p_c$ . Fig. 8 shows the course of edge orientation for the points  $p_i$  located on the line segment  $\mathcal{L}(p_a, p_d)$ . The horizontal axis is for the points on the line segment and the vertical axis for the orientations. Furthermore, the orientation  $\phi$  of the line segment  $\mathcal{L}(p_a, p_d)$  is depicted, which is of course independent of the points on the line. For the points of the line segment  $\mathcal{L}(p_b, p_c)$  we obtain small deviation values between the edge orientations and the line orientation. On the other hand, there is large variance in the set of deviation values coming from edge orientations of the points of the line segments  $\mathcal{L}(p_a, p_b)$  or  $\mathcal{L}(p_c, p_d)$ .

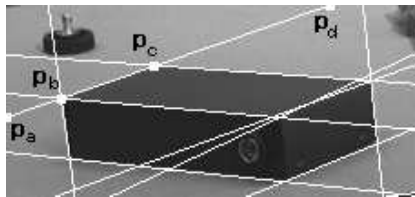


FIG. 7: Example line with characteristic points  $\{p_a, p_b, p_c, p_d\}$ , defined by intersection with other lines and with the image border. Just the segment between  $p_b$  and  $p_c$  is relevant for the boundary.

For verifying a line segment we evaluate the deviation between the orientation of the line and the orientations of the gray value edges of all points on the line segment.

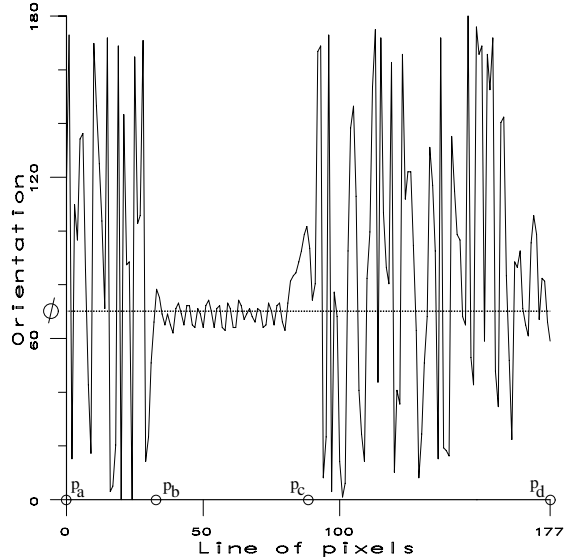


FIG. 8: Course of edge orientation of points along the line  $\mathcal{L}(p_a, p_d)$ , and orientation  $\phi$  of this line. Small deviation within the relevant segment  $\mathcal{L}(p_b, p_c)$ , and large deviations within the other two segments.

**Definition 4 (Orientation-deviation related to a line segment)** *The orientation-deviation between orientation  $\phi$  of a line and the orientations of all edges on a segment  $\mathcal{L}(p_1, p_N)$  of the line is defined by*

$$D_{LE}(\phi, \mathcal{L}(p_1, p_N)) := \frac{1}{N} \sum_{i=1}^N D_O(\phi, O(p_i)) \quad (5)$$

$$D_O(\phi, O(p_i)) := \frac{\min\{|\phi - O(p_i)|, |\phi - O(p_i) + 180^\circ|, |\phi - O(p_i) - 180^\circ|\}}{90^\circ} \quad (6)$$

The minimization involved in Eq. (6) is due to the restriction of edge and line orientation in the angle interval  $[0^\circ, \dots, 180^\circ]$  respectively. As an example of application, the deviation between a line orientation  $\phi = 0^\circ$  and an edge orientation  $O(p_i) = 180^\circ$  must be defined to be zero. Furthermore, a normalization factor is introduced to restrict the deviation values in the real unit interval. The orientation-deviation related to the line segments  $\mathcal{L}(p_a, p_b)$ ,  $\mathcal{L}(p_b, p_c)$ , and  $\mathcal{L}(p_c, p_d)$  in Fig. 7 is shown in Fig. 9. For line segment  $\mathcal{L}(p_b, p_c)$  it is minimal, as expected, because this line segment originates from the actual boundary of the target object.

Based on the definition for orientation-deviation we can formally introduce the line/edge orientation-consensus between the orientation of a line and the orientations of all edges on a segment of the line.

**Assumption 2 (Line/edge orientation-consensus for a line segment)** *Let  $\delta_3$  be the permissible orientation-deviation in the sense of a necessary geometric/photometric consensus. The line/edge orientation-consensus holds between the orientation  $\phi$  of a line and the orientations  $O(p_i)$  of all edges on a segment  $\mathcal{L}(p_1, p_N)$  of the line, if*

$$D_{LE}(\phi, \mathcal{L}(p_1, p_N)) \leq \delta_3 \quad (7)$$



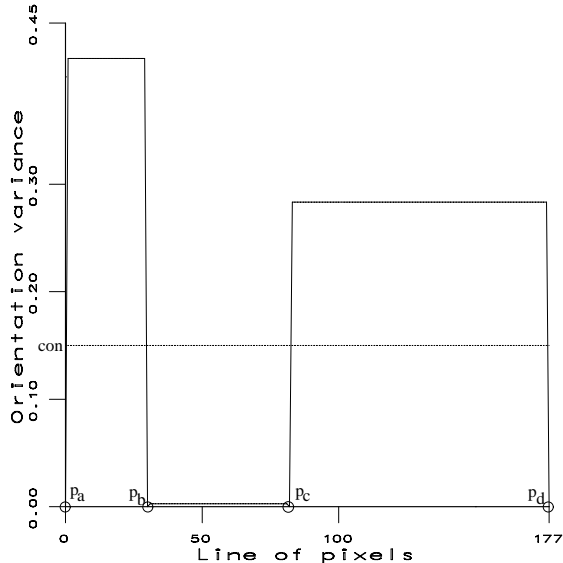


FIG. 9: Mean variance of the edge orientations for three line segments  $\mathcal{L}(p_a, p_b)$ ,  $\mathcal{L}(p_b, p_c)$ ,  $\mathcal{L}(p_c, p_d)$  related to line orientation  $\phi$ .

For example, Fig. 9 shows that the line/edge orientation-consensus just holds for the line segment  $\mathcal{L}(p_b, p_c)$ , if we apply a consensus threshold  $\delta_3 = 0.15$ . For the extraction of object boundaries we will generally assume that the line/edge orientation-consensus just holds for the line segments, which are relevant for a boundary.

### 2.3 Junction-consensus between convergences and corners

The geometric *line* feature and the photometric *edge* feature are one-dimensional in nature. A further sophisticated consensus criterion can be defined on the basis of two-dimensional image structures. In the projected object boundary usually two or more lines meet at a common point (see points  $p_b$  and  $p_c$  in Fig. 7). A collection of non-parallel image line segments meeting at a common point is designated as a *line convergence*. At the convergence point a *gray value corner* should be detected in the image. Generally, gray value corners are located at the curvature extrema along edge sequences. A review of several corner detectors was presented in [32], however we used the recently published SUSAN operator [33]. Exemplary, Fig. 10 shows a set of gray value corners with especially the corners at the points  $p_b$  and  $p_c$  included.

According to that we must consider a consensus between the geometric *convergence* feature and the photometric *corner* feature. Common attributes are needed for characterizing a junction of lines and a junction of edge sequences. We define a  $(\#M)$ -*junction* to consist of  $M$  converging lines resp. edge sequences. A  $(\#M)$ -*junction of lines* can be characterized by the position  $p_{cv}$  of the convergence point and the orientations  $\mathcal{A} := (\alpha_1, \dots, \alpha_M)$  of the converging lines related to the horizontal axis. Similarly, a  $(\#M)$ -*junction of edge sequences* is characterized by the position  $p_{cr}$  of the corner point and the orientations  $\mathcal{B} := (\beta_1, \dots, \beta_M)$  of the converging edge sequences against the horizontal axis.



FIG. 10: The SUSAN operator has detected a set of gray value corners shown as white squares. We find all those gray value corners included which are characteristic for the three-dimensional object boundary.

**Definition 5 (Junction-deviation related to a line convergence)** *The junction-deviation between a  $(\#M)$ -junction of lines with orientations  $\mathcal{A}$  at a convergence point  $p_{cv}$  and a  $(\#M)$ -junction of edge sequences with orientations  $\mathcal{B}$  at a corner point  $p_{cr}$  is defined by*

$$D_{CC}(p_{cv}, p_{cr}, \mathcal{A}, \mathcal{B}) := \lambda_1 * D_{JP}(p_{cv}, p_{cr}) + \lambda_2 * D_{JO}(\mathcal{A}, \mathcal{B}) \quad (8)$$

$$D_{JP}(p_{cv}, p_{cr}) := \frac{\|p_{cv} - p_{cr}\|}{I_d} \quad (9)$$

$$D_{JO}(\mathcal{A}, \mathcal{B}) := \frac{1}{180^\circ * M} \sum_{i=1}^M \min\{|\alpha_i - \beta_i|, |\alpha_i - \beta_i + 360^\circ|, |\alpha_i - \beta_i - 360^\circ|\} \quad (10)$$

Eq. (8) combines two components of junction-deviation with the factors  $\lambda_1$  and  $\lambda_2$ , which are used to weight each part. The first component (Eq. (9)) evaluates the euclidean distance between convergence and corner point. The second component (Eq. (10)) computes the deviation between the orientation of a line and of the corresponding edge sequence, and this is done for all corresponding pairs in order to compute a mean value. Both components and the final outcome of Eq. (8) are normalized in the real unit interval. Based on this definition we formally introduce a convergence/corner junction-consensus.

**Assumption 3 (Convergence/corner junction-consensus)** *Let  $\delta_4$  be the permissible junction-deviation in the sense of a necessary geometric/photometric consensus. The convergence/corner junction-consensus (CCJC) holds between a  $(\#M)$ -junction of lines and a  $(\#M)$ -junction of edge sequences, if*

$$D_{CC}(p_{cv}, p_{cr}, \mathcal{A}, \mathcal{B}) \leq \delta_4 \quad (11)$$

This convergence/corner junction-consensus will be used as a criterion to evaluate whether a junction belongs to the boundary of a target object. For illustration it is applied to three junctions designated in Fig. 11 by the indices 1, 2, 3. It is remarkable that junctions 1 and 2 belong to the boundary of the target object but junction 3 doesn't. The white squares show the convergence points  $p_{cv1}, p_{cv2}, p_{cv3}$ . They are determined by intersection of straight lines extracted via OHT (see also Fig. 6 (bottom)). The black squares show the corner points  $p_{cr1}, p_{cr2}, p_{cr3}$ . They have been selected from the whole set of corner points in Fig. 10 based on nearest neighborhood to the convergence points. Related to Eq. (9) we realize that all three convergence points have corresponding corner points

in close neighborhood. I.e. for the three junctions 1, 2, 3 we obtain position deviations (normalized by the image diagonal  $I_d$ ) of about 0.009, 0.006, 0.008 respectively. According to that, a small position deviation is only a necessary but not a sufficient criterion to classify junction 3 unlike to 1 and 2. Therefore, the junctions must be characterized in more detail, which is considered in Eq. (10).

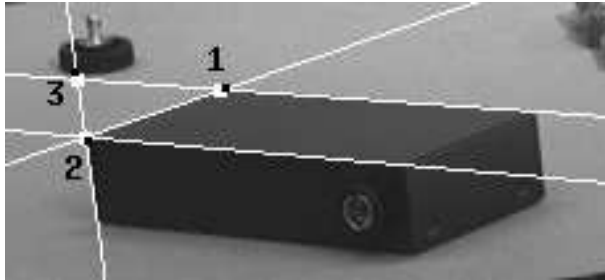


FIG. 11: A subset of three gray value corners is selected. They are located in close neighborhood to three convergence points of relevant boundary lines respectively. Just the corners at junctions 1 and 2 are relevant for boundary extraction but not the corner at junction 3.

A steerable wedge filter, adopted from [34], has been applied at the convergence points  $p_{cv1}, p_{cv2}, p_{cv3}$  in order to locally characterize the gray value structure. We prefer convergence points instead of the neighboring corner points, because the convergence points arise from line detection, which is more robust than corner detection. A wedge is rotating in discrete steps around a convergence point and at each step the mean gray value within the wedge mask is computed. E.g. the wedge started in horizontal direction pointing to the right and then rotated counter-clockwise in increments of  $4^\circ$  angle degrees. According to the steerability property of this filter we compute the filter response at the basic orientations, and approximate filter responses in between two basic orientations (if necessary). This gives a one-dimensional course of smoothed gray values around the convergence point. The first derivative of a one-dimensional Gaussian is applied to this course and the magnitude computed from it. We obtain for each discrete orientation a *significance measurement* for the existence of an edge sequence flowing with that orientation into the convergence point. As a result, a curve of filter responses is obtained, which characterizes the gray level structure around the convergence point.

These curves are shown for the junctions 1, 2, and 3 in Fig. 12, Fig. 13, and Fig. 14 respectively. The curve in Fig. 12 shows two local maxima (near to  $0^\circ$  respective  $360^\circ$  and near to  $200^\circ$ ) indicating a (#2)-junction. The curve in Fig. 13 shows three local maxima (near to  $20^\circ$ ,  $290^\circ$ , and  $360^\circ$ ) indicating a (#3)-junction. The curve in Fig. 14 shows two local maxima (near to  $20^\circ$  and  $150^\circ$ ) indicating a (#2)-junction. The vertical dotted lines in each figure indicate the orientation of the image lines (extracted by Hough transformation), which converge at the three junctions respectively. We clearly observe that in Fig. 12 and Fig. 13 the local maxima are located near to the orientations of the converging lines. However, in Fig. 14 the positions of the curve maxima and of the dotted lines differ significantly because junction 3 doesn't belong to the object boundary. By applying the formula in Eq. (10) we compute about 0.04, 0.03, 0.83 for the three junctions 1, 2, 3 respectively. Based on a threshold we can easily conclude that the convergence/corner junction-consensus holds for junctions 1 and 2, but not for junction 3.

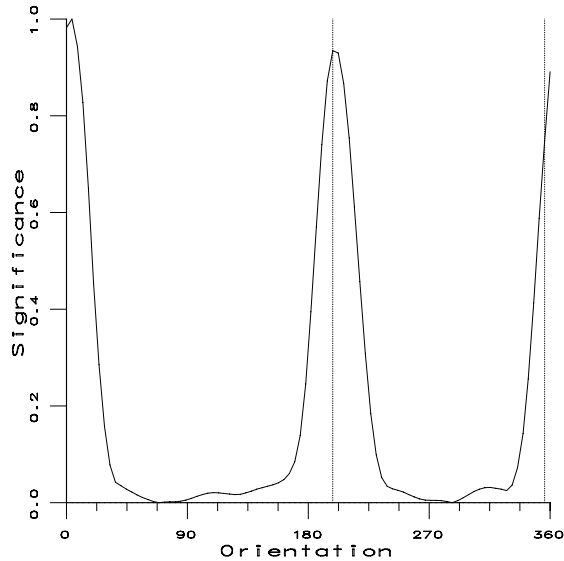


FIG. 12: Response of a counter-clockwise rotating wedge filter applied at junction 1. The two local maxima indicate a (#2)-junction, i.e. two converging edge sequences. The orientations of the converging edge sequences (maxima of the curve) are similar to the orientations of two converging lines (denoted by the positions of the vertical lines). The convergence/corner junction-consensus holds for junction 1.

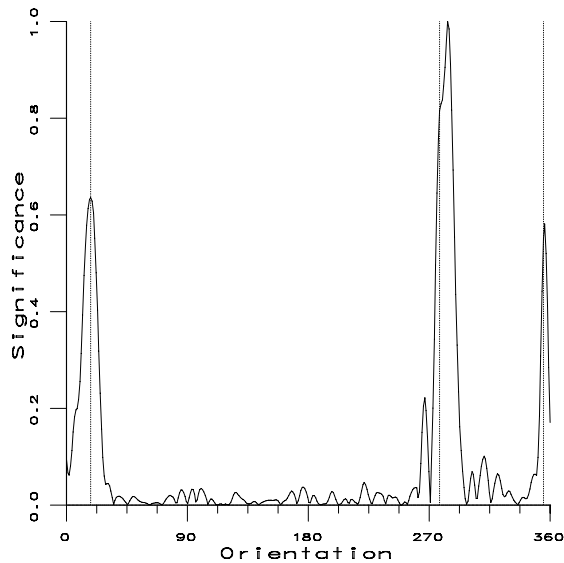


FIG. 13: Response of a counter-clockwise rotating wedge filter applied at junction 2. The three local maxima indicate a (#3)-junction, i.e. three converging edge sequences. The maxima of the curve are near the positions of the vertical lines. The convergence/corner junction-consensus holds for junction 2.

The line/edge orientation-consensus and the convergence/corner junction-consensus can generally be assumed for all scenes containing approximate polyhedral objects. In Sections 3 and 4 the evaluation of the junction-consensus according to Eq. (8) is combined with the evaluation of the orientation-consensus according to Eq. (5). This measure will be

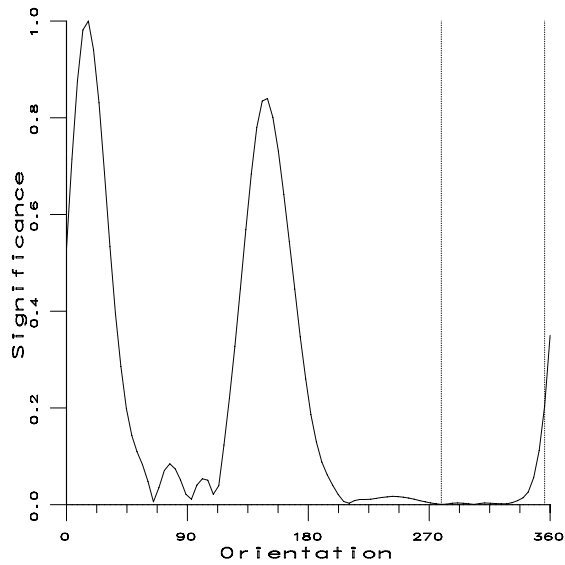


FIG. 14: Response of a counter-clockwise rotating wedge filter applied at junction 3. The two local maxima indicate a (#2)-junction. The orientations of the converging edge sequences are dissimilar to those of the converging lines. The convergence/corner junction-consensus does not hold for junction 3.

used in combination with regularity features of object shapes to define the relevance of certain line segments for the boundary description.

### 3 Consensus-based structural level grouping

The orientation-selective Hough transformation (OHT), the line/edge orientation-consensus (LEOC) and the convergence/corner junction-consensus (CCJC) are the basis for detecting high-level geometric structures in the image. Additionally, we make dominant use of certain regularity features, which are inherent in man-made 3D objects and are quasi-invariant under projective transformation. Approximate parallel or right-angled line segments, or approximate reflection-symmetric or translation-symmetric polylines are considered in a sophisticated search strategy for detecting organizations of line segments. This section focuses on the extraction of polygons originating from the faces or the silhouettes of approximate polyhedral 3D objects. Related to the aspect of extracting polygons our approach is similar to Havaldar et al. [35], who also extract closed figures, e.g. by discovering approximate reflection-symmetric polylines (they call it *skewed symmetries between super segments*). The principal distinction to our work is that we strive for an integration of grouping cues from this structural level with cues from other levels, i.e. from signal level and assembly level.

#### 3.1 Configuration of Hough peaks for approximate parallel lines

It is well-known for the projective transformation of an ideal pinhole camera that an image point  $(x_1, x_2)$  is computed from a 3D scene point  $(y_1, y_2, y_3)$  by

$$x_1 := f * \frac{y_1}{y_3} \quad , \quad x_2 := f * \frac{y_2}{y_3} \quad , \quad (12)$$

where  $f$  is the distance between the lens center and the projection plane. Furthermore, according to Eq. (12) it is obvious that parallel 3D lines are no longer parallel after projective transformation to the image (except for lines parallel to the projection plane). Fortunately, for certain imaging conditions the parallelism is *quasi-invariant* under projective transformation. To get an impression for that we describe the imaging condition for taking the picture in Fig. 15. The distance between the camera and the target object was about  $1000mm$ , and the lens of the objective was of  $12mm$  focal length. In that case the deviation from parallelism, which depends on object orientation, is at most  $15^\circ$  angle degrees. We will formulate the parallelism quasi-invariant and relate it to the configuration of Hough peaks.

**Definition 6 (Approximate parallel lines)** *Let  $\delta_5$  be the permissible deviation from parallelism, i.e. maximal deviation from exact regularity. Two image lines with values  $\phi_1$  and  $\phi_2$  of the angle parameter are approximate parallel, if*

$$D_O(\phi_1, \phi_2) \leq \delta_5 \quad (13)$$

**Assumption 4 (Quasi-invariance of parallelism)** *The parallelism is a quasi-invariant, if parallel lines in 3D are approximate parallel after projective transformation. For such imaging conditions, parallel lines in 3D occur as peaks in the Hough image being located within a horizontal stripe of height  $\delta_5$ .*



FIG. 15: Arrangement of objects in a scene of electrical scrap. The black dummy box is our target object for the purpose of demonstration. It is located in a complex environment, is partially occluded, and has a protruding socket.

The peaks in a Hough stripe correlate with approximate parallel lines in the gray value image. Fig. 16 shows the Hough image obtained from the gray value image in Fig. 15 after binarization (edge detection) and application of the OHT.

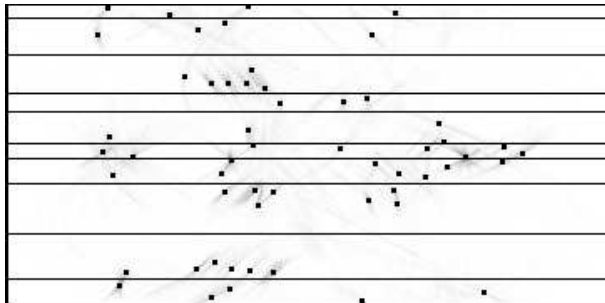


FIG. 16: Hough image obtained after edge detection in the image of Fig. 15. A set 55 most maximal peaks is marked by black dots. They have been organized in 10 clusters (horizontal stripes) using the ISODATA clustering algorithm.

The Hough image has been edited with black squares and horizontal lines, which mark a set of 55 local maxima organized in 10 horizontal stripes. The local maxima are obtained using the approach mentioned in section 2.2. For grouping the peak positions we just take parameter  $\phi$  into account and use the distance function  $D_O$  from Eq. (6). According to that, angles near to  $0^\circ$  can be grouped with angles near to  $180^\circ$ . A procedure similar to the error-based *ISODATA clustering algorithm* can be applied but taking the modified distance function into account (see Schalkoff et al. [36], pp. 109–125). Initially, the algorithm groups vectors (in this application, simply scalars) by using the standard *K-means* method. Then, clusters exhibiting large variances are split in two, and clusters

that are too close together are merged. Next, K-means is reiterated taking the new clusters into account. This sequence is repeated until no more clusters are split or merged. The merging/splitting parameters are taken in agreement with the pre-specified  $\delta_5$  from Assumption 4.

For example, Fig. 17 shows the set of approximate parallel lines specified by the Hough peaks in the fourth stripe, and Fig. 18 shows it for the eighth stripe of Hough peaks. Under these lines we find candidates for describing the boundary of the dummy box. In the next subsections the horizontal clusters of Hough peaks are used in combination with the LEOC and CCJC principles for extracting the faces or silhouettes of objects (or merely approximate faces or silhouettes) from the images.

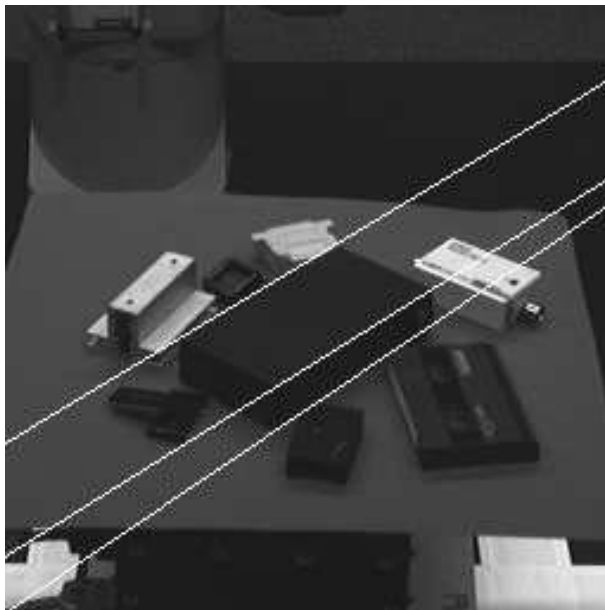


FIG. 17: The set of approximate parallel lines specified by the Hough peaks in the fourth stripe of Fig. 16.

### 3.2 Extraction of regular quadrangles

In a multitude of man-made objects the faces or the silhouettes can be approximated by squares, rectangles, or trapezoids. The projective transformation of these shapes yields approximations of squares, rhombuses, rectangles, parallelograms, or trapezoids. A generic procedure will be presented for extracting from the image these specific quadrangles (The next subsection, after this one, presents a procedure for extracting more general polygons). Under the constraint of clustered Hough peaks we exhaustively look for quadruples of Hough peaks, extract the four line segments by line intersection respectively, evaluate the LEOC and CCJC principles, and determine deviations from a certain standard form. Finally, for each quadrangle all evaluations are combined, which results in a saliency value including both photometric and geometric aspects.

We have introduced the orientation-deviation  $D_{LE}$  related to a line segment in Eq. (5) and the junction-deviation  $D_{CC}$  related to a convergence of line segments in Eq. (8). In order to extend the LEOC and CCJC principles to quadrangles we simply average these



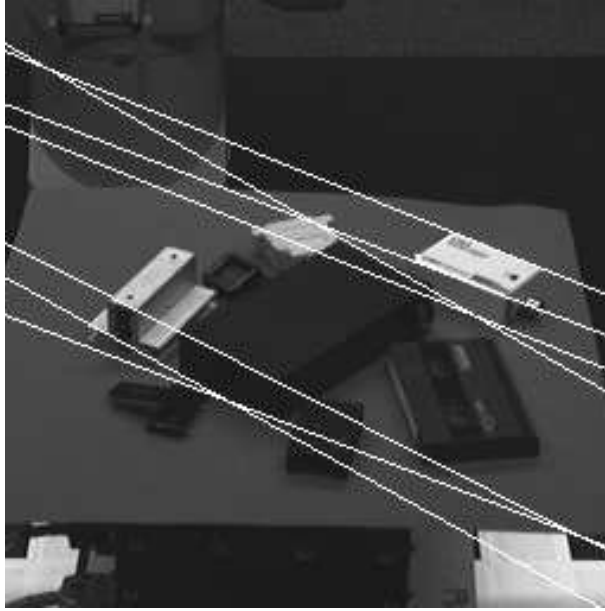


FIG. 18: The set of approximate parallel lines specified by the Hough peaks in the eighth stripe of Fig. 16.

values for the four segments respective for the four convergences.

$$D_{LE\_QD} := \sum_{i=1}^4 D_{LE\_i}, \quad D_{CC\_QD} := \sum_{i=1}^4 D_{CC\_i} \quad (14)$$

For convenience we have omitted the parameters and simply have introduced an index for the line segments involved in a quadrangle. The resulting functions  $D_{LE\_QD}$  and  $D_{CC\_QD}$  can be used in combination to define a geometric/photometric consensus for quadrangles.

**Assumption 5 (Geometric/photometric consensus for a quadrangle)** *The necessary geometric/photometric consensus for a quadrangle is specified by parameter  $\delta_6$ . The geometric/photometric consensus for a quadrangle holds, if*

$$(D_{LE\_QD} + D_{CC\_QD}) \leq \delta_6 \quad (15)$$

To consider the pure geometric aspect we define the deviation of a quadrangle from certain standard forms. For the sequence of four line segments of a quadrangle, let  $\mathcal{S} := (s_1, s_2, s_3, s_4)$  be the lengths,  $\mathcal{G} := (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$  be the inner angles of two successive segments, and  $\mathcal{F} := (\phi_1, \phi_2, \phi_3, \phi_4)$  be the orientation angles of the polar form representations.

**Definition 7 (Rectangle-deviation, parallelogram-deviation, square-deviation, rhombus-deviation, trapezoid-deviation)**

*The rectangle-deviation of a quadrangle is defined by*

$$D_{RC}(\mathcal{G}) := \frac{1}{4 * 360^\circ} * \sum_{i=1}^4 |\gamma_i - 90^\circ| \quad (16)$$

The parallelogram-deviation of a quadrangle is defined by

$$D_{PA}(\mathcal{G}) := \frac{1}{2 * 360^\circ} * (|\gamma_1 - \gamma_3| + |\gamma_2 - \gamma_4|) \quad (17)$$

The square-deviation of a quadrangle is defined by

$$D_{SQ}(\mathcal{G}, \mathcal{S}) := \frac{1}{2} * (D_{RC}(\mathcal{G}) + V_{SL}(\mathcal{S})) \quad (18)$$

with the normalized length variance  $V_{SL}(\mathcal{S})$  of the four line segments.

The rhombus-deviation of a quadrangle is defined by

$$D_{RH}(\mathcal{G}, \mathcal{S}) := \frac{1}{2} * (D_{PA}(\mathcal{G}) + V_{SL}(\mathcal{S})) \quad (19)$$

The trapezoid-deviation of a quadrangle is defined by

$$D_{TR}(\mathcal{F}) := \min\{D_O(\phi_1, \phi_3), D_O(\phi_2, \phi_4)\} \quad (20)$$

Normalization factors are chosen such that the possible values of each function fall in the unit interval respectively. For the rectangle-deviation the mean deviation from right-angles is computed for the inner angles of the quadrangle. For the parallelogram-deviation the mean difference between diagonally opposite inner angles is computed. The square-deviation and rhombus-deviation are based on the first two respectively and additionally include the variance of the lengths of line segments. The trapezoid-deviation is based on Eq. (6) and computes for the two pairs of diagonally opposite line segments the minimum of deviation from parallelism.

The features related to the geometric/photometric consensus for quadrangles, and the feature related to the geometric deviation of quadrangles from specific shapes must be combined to give measures of conspicuity of certain shapes in an image.

**Definition 8 (Saliency of specific quadrangles)** *The saliency of a specific quadrangle is defined by*

$$S_{SP\_QD} := 1 - (\lambda_1 * D_{LE\_QD} + \lambda_2 * D_{CC\_QD} + \lambda_3 * D_{SP\_QD}) \quad (21)$$

The specific quadrangle can be an approximate rectangle, parallelogram, square, rhombus, or trapezoid, and for these cases the generic function symbol  $D_{SP\_QD}$  must be replaced by  $D_{RC}$ ,  $D_{PA}$ ,  $D_{SQ}$ ,  $D_{RH}$ , or  $D_{TR}$ .

## Generic procedure for the extraction of specific quadrangles:

1. For each pair of cluster stripes in the set of Hough peaks:
  - 1.1. For each pair of Hough peaks in the first stripe of the pair:
    - 1.1.1. For each pair of Hough peaks in the second stripe of the pair:
      - 1.1.1.1. Intersect the lines specified by the four Hough peaks and construct the quadrangle.
      - 1.1.1.2. Compute the mean line/edge orientation-deviation using function  $D_{LE\_QD}$ .
      - 1.1.1.3. Compute the mean convergence/corner junction-deviation using function  $D_{CC\_QD}$ .
      - 1.1.1.4. Compute the deviation from the specific quadrangle using function  $D_{SP\_QD}$ .
      - 1.1.1.5. Compute the saliency value by combining the above results according to Eq. (21).
2. Bring the specific quadrangles into order according to decreasing saliency values.

The generic procedure works for all types of specific quadrangles, mentioned above, except for trapezoids. For the extraction of trapezoids the algorithm is modified, such that it iterates over single cluster stripes, and selects all combinations of two Hough peaks in each stripe respectively, and takes the third and fourth Hough peaks from any other cluster stripes. Thus it will be considered that a trapezoid just consists of one pair of parallel line segments, instead of two such pairs of the other specific quadrangles.

These procedures have been applied to complicated scenes of electrical scrap in order to draw conclusions concerning the usefulness of the principles introduced above. The goal of the following experiments was to extract from the images specific quadrangles, which describe the faces or silhouettes of objects. First, we have applied the procedure to the image in Fig. 15 with the intention of extracting approximate parallelograms. In the saliency measure the weighting factors  $\lambda_1, \lambda_2, \lambda_3$  are set equal to 0.33. Fig. 19 shows exemplary a set of 65 parallelograms, which are best related to the saliency measure. The silhouette boundary of the dummy box (see Fig. 20) is included as number nine. The decreasing course of saliency values for the parallelograms is depicted in Fig. 21. Second, the procedure is used to extract approximate rectangles with the intention of extracting the electronic board of a computer interior in Fig. 22. The best set of 10 approximate rectangles are outlined in black color, including the relevant one of the board. Third, the procedure has been applied to extract approximate rhombuses with the goal of locating the electronic board in the image of Fig. 23. The best set of 4 approximate rhombuses are outlined in white color, including the relevant one. Finally, the extraction of approximate trapezoids is shown in Fig. 24.

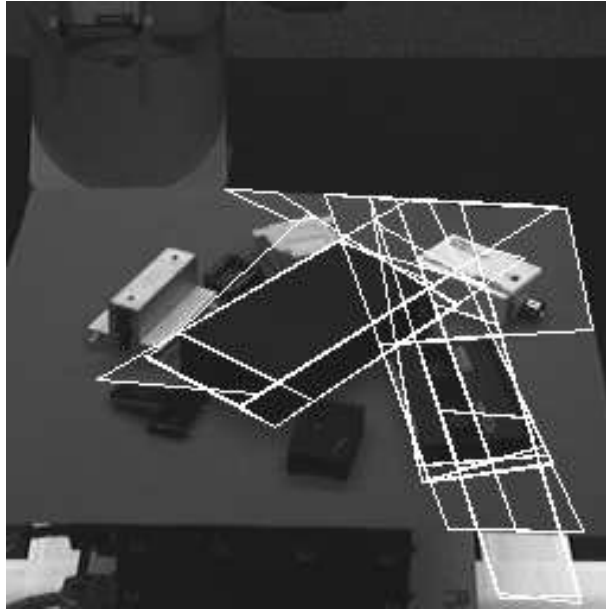


FIG. 19: Based on a saliency measure a subset of most conspicuous, approximate parallelograms have been extracted.



FIG. 20: The silhouette boundary of the dummy box is included in the set of most conspicuous, approximate parallelograms.

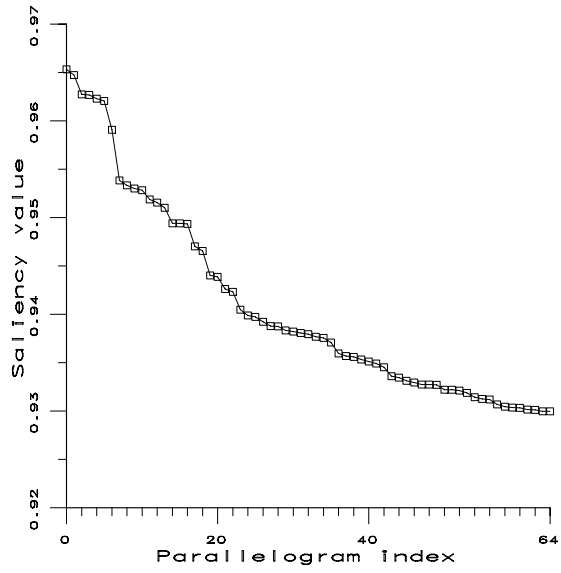


FIG. 21: Decreasing course of saliency values for the subset of extracted parallelograms in the image in Fig. 19.

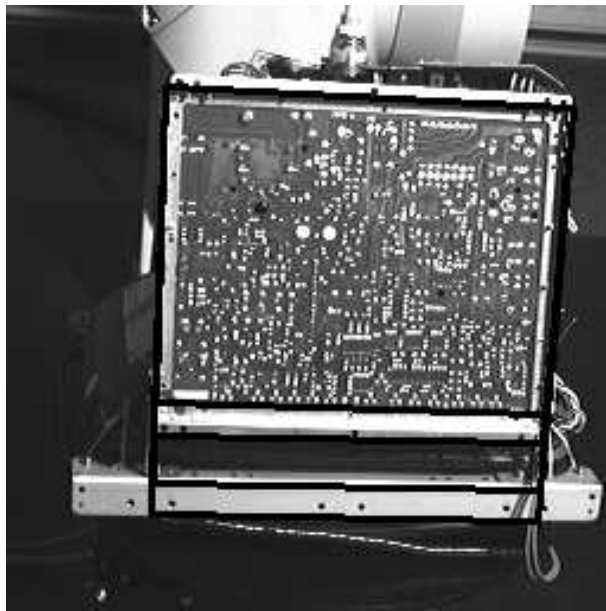


FIG. 22: Electronic board of a computer interior and an extracted subset of approximate rectangles. One of these rectangles represents the boundary of the board.

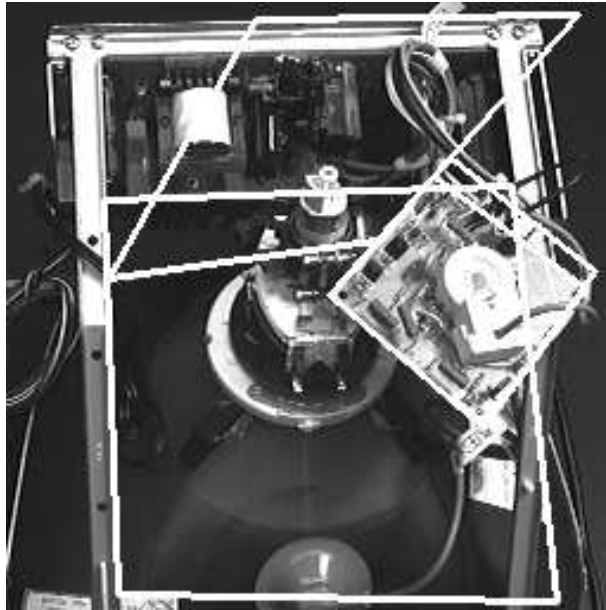


FIG. 23: Computer interior containing an electronic board. One of the extracted approximate rhombuses represents the boundary of the board.

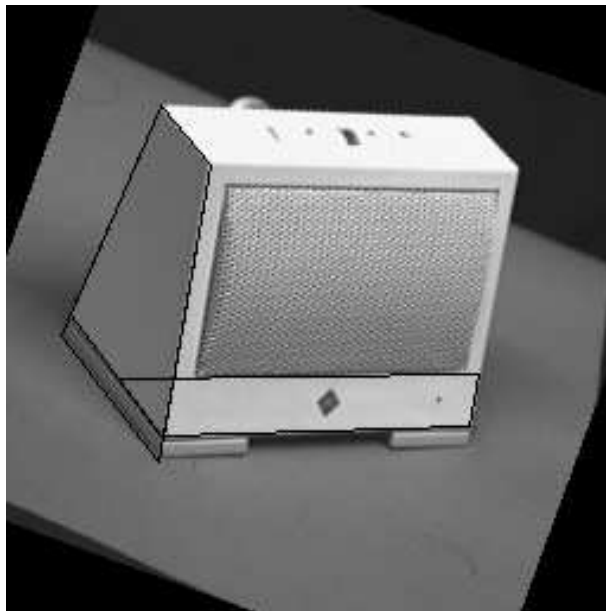


FIG. 24: Image of a loudspeaker and a subset of extracted approximate trapezoids. One of these represents a side-face of the loudspeaker.

### 3.3 Extraction of regular polygons

The measures  $D_{LE}$  and  $D_{CC}$  involved in the geometric/photometric consensus of single line segments can easily be extended to polygons of  $K$  line segments.

$$D_{LE\_PG} := \sum_{i=1}^K D_{LE\_i}, \quad D_{CC\_PG} := \sum_{i=1}^K D_{CC\_i} \quad (22)$$

**Assumption 6 (Geometric/photometric consensus for a polygon)** *The necessary geometric/ photometric consensus for a polygon is specified by parameter  $\delta_7$ . The geometric/photometric consensus for a polygon holds, if*

$$(D_{LE\_PG} + D_{CC\_PG}) \leq \delta_7 \quad (23)$$

These consensus features must be combined with pure geometric features of the polygon. By considering specific polygon regularities a measurement of conspicuity is obtained, i.e. a saliency value for the polygon. More complex regularities are interesting than the simple ones involved in specific quadrangles. For polygons with arbitrary segment number we define three types of regularities, which are general in the sense that they typically appear in scenes of man-made objects. Two types of regularities are based on symmetries between polylines, i.e. the reflection-symmetry and the translation-symmetry (defined later on). These appear exemplary in the image of a computer monitor as side and top face (see in Fig. 25 the two polygons outlined with white color). The third type of regularity is a right-angle, in which two successive line segments of a polygon are right-angled respectively. For example, Fig. 26 shows an electronic board with approximate right-angled shape.



FIG. 25: Computer monitor with two types of regularities of the polygonal faces. The side face is approximate reflection-symmetric, and the top face is approximate translation-symmetric.

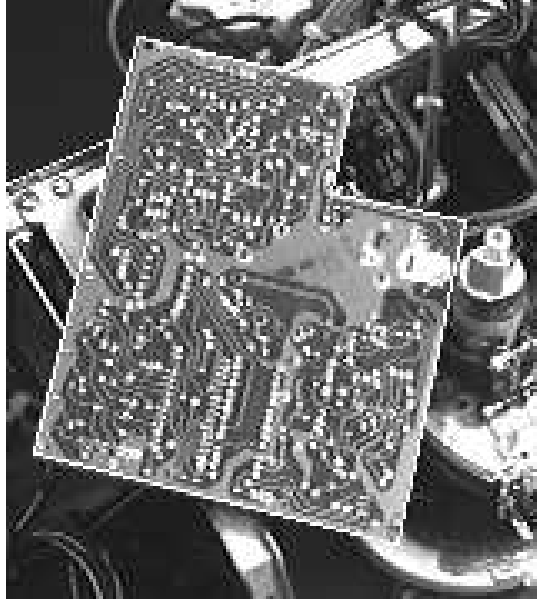


FIG. 26: Computer interior containing an electronic board, whose boundary is represented by a hexagon. The line convergences of the board hexagon are approximate right-angles.

A *polyline*, involved in the definition of symmetries, is a non-closed and non-branching sequence of connected line segments. We construct for each polygon a pair of non-overlapping polylines with equal numbers of line segments. A polygon with an odd number of line segments is the union of two polylines and one single line segment (see in Fig. 27, top left and right). For a polygon with an even number of line segments we distinguish two cases. First, the polygon can be the union of two polylines, i.e. they meet at two polygon junctions (see in Fig. 27, middle). Second, the polygon can be the union of two polylines and of two single line segments located at the end of each polyline respectively (see Fig. 27, bottom).

Let  $\mathcal{G} := (\gamma_1, \dots, \gamma_K)$  be the ordered sequence of inner angles of the polygon. A candidate pair of polylines is represented by  $\mathcal{G}^1 := (\gamma_1^1, \dots, \gamma_k^1)$ , i.e. the sequence of inner angles related to the first polyline, and by  $\mathcal{G}^2 := (\gamma_1^2, \dots, \gamma_k^2)$ , i.e. the opposite (corresponding) inner angles related to the second polyline. Each  $\gamma_i^1$  or  $\gamma_i^2$  is contained in  $\mathcal{G}$ . Included in  $\mathcal{G}^1$  and  $\mathcal{G}^2$  are the inner angles at the end of the polylines, where one polyline meets the other one or meets a single line segment. The set of all pairs  $(\mathcal{G}^1, \mathcal{G}^2)$  of angle sequences, which describe all candidate pairs of polylines of a polygon, is designated by  $\mathcal{G}_\Sigma$ . For example, Fig. 27 (bottom) shows for a candidate pair of polylines the two corresponding sequences of inner polygon angles.

**Definition 9 (Reflection-symmetric polygon)** *A polygon is reflection-symmetric, if a pair of polylines exists with sequences  $\mathcal{G}^1$  and  $\mathcal{G}^2$  of inner angles such that  $d_{rs}(\gamma_i^1, \gamma_i^2) = 0^\circ$  for each tuple  $(\gamma_i^1, \gamma_i^2), i \in \{1, \dots, k\}$ ,*

$$d_{rs}(\gamma_i^1, \gamma_i^2) := |\gamma_i^1 - \gamma_i^2| \quad (24)$$

Fig. 27 shows three reflection-symmetric polygons (left column) and the relevant configuration of polylines and single line segments (right column). Obviously, for all three



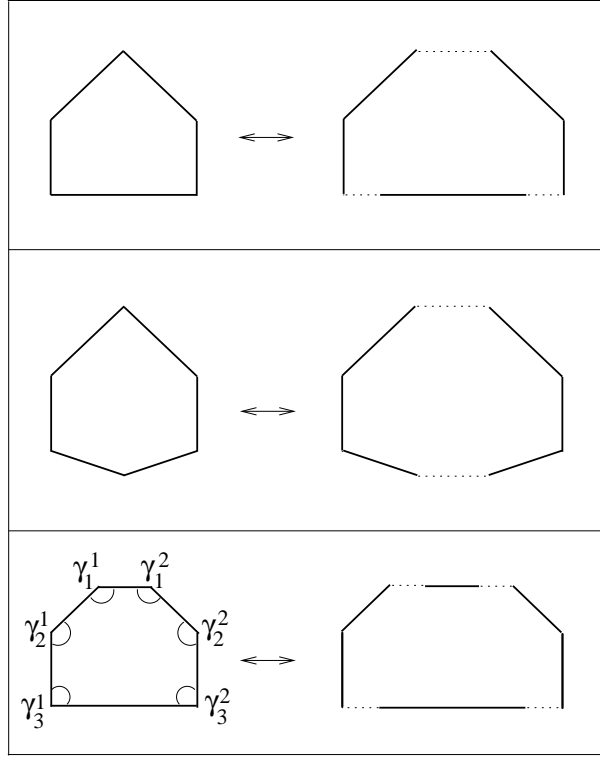


FIG. 27: Organization of reflection-symmetric polygons by two polylines with equal number of line segments and up to two single line segments.

polygons there exist a vertical axis of reflection for mapping one polyline onto the other. Exemplary, in the bottom polygon we realize:  $\gamma_1^1 = \gamma_1^2$ ,  $\gamma_2^1 = \gamma_2^2$ ,  $\gamma_3^1 = \gamma_3^2$ .

**Definition 10 (Translation-symmetric polygon)** *A polygon is translation-symmetric, if a pair of polylines exists with sequences  $\mathcal{G}^1$  and  $\mathcal{G}^2$  of inner angles such that  $d_{ts}(\gamma_i^1, \gamma_i^2) = 0^\circ$  for each tuple  $(\gamma_i^1, \gamma_i^2), i \in \{1, \dots, k\}$ ,*

$$d_{ts}(\gamma_i^1, \gamma_i^2) := \begin{cases} |\gamma_i^1 - (360^\circ - \gamma_i^2)| & : i \in \{2, \dots, k-1\} \\ |\gamma_i^1 - (180^\circ - \gamma_i^2)| & : i \in \{1, k\} \end{cases} \quad (25)$$

For the translation-symmetry it is plausible to match the inner polygon angles ( $\gamma_i^1$ ) related to the first polyline with the corresponding exterior angles ( $360^\circ - \gamma_i^2$ ) related to the second polyline. The corresponding angles at the end of the two polylines respectively must be matched modulo  $180^\circ$  angle degrees. Fig. 28 shows a hexagon with translation-symmetry and the relevant pair of polylines. We can imagine a translation vector for mapping one polyline onto the other. The property of translation-symmetry is easily realized by:  $\gamma_1^1 = (180^\circ - \gamma_1^2)$ ,  $\gamma_2^1 = (360^\circ - \gamma_2^2)$ ,  $\gamma_3^1 = (180^\circ - \gamma_3^2)$ .

For a polygon the property of reflection-symmetry or translation-symmetry is examined by making an iteration over all possible pairs of polylines and perhaps finding one, for which the proposition in Definition 9 respective 10 holds. Otherwise, the polygon is not regular related to reflection- or translation-symmetry. Fig. 29 shows two polygons with inappropriate pairs of polylines, though the reflection-symmetry respective the translation-symmetry has already been realized in Fig. 27 (bottom) and Fig. 28.

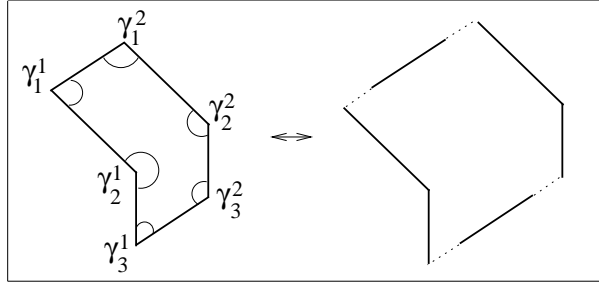


FIG. 28: Organization of a translation-symmetric hexagon by two polylines and two single line segments.

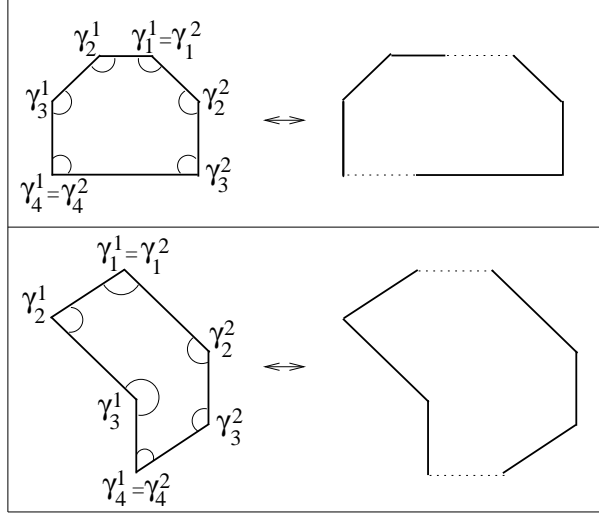


FIG. 29: Polygons with inappropriate organization of pairs of polilines, respectively. The reflection-symmetry respective the translation-symmetry is examind based on the alternative organizations in Fig. 27 (bottom) and Fig. 28.

For a polygon the deviation from reflection-symmetry respectively translation-symmetry is defined by matching for all candidate pairs of polylines the two sequences of inner angles and taking the minimum value.

**Definition 11 (Deviation from reflection-symmetry or translation-symmetry)**

*For a polygon the deviation from reflection-symmetry is*

$$D_{RS} := \min_{\mathcal{G}_\Sigma} \left\{ \frac{1}{k * 360^\circ} \sum_{i=1}^k d_{rs}(\gamma_i^1, \gamma_i^2) \right\} \quad (26)$$

*For a polygon the deviation from translation-symmetry is*

$$D_{TS} := \min_{\mathcal{G}_\Sigma} \left\{ \frac{1}{k * 360^\circ} \sum_{i=1}^k d_{ts}(\gamma_i^1, \gamma_i^2) \right\} \quad (27)$$

For the reflection-symmetric polygons in Fig. 27 the equation  $D_{RS} = 0$  is obtained and for the translation-symmetric polygon in Fig. 28 the equation  $D_{TS} = 0$ . However, for the left face of the computer monitor in Fig. 25 we compute  $D_{RS} \approx 0.064$ , which indicates an

approximate reflection-symmetry. For the top face of the computer monitor we compute  $D_{TS} \approx 0.015$ , which indicates an approximate translation-symmetry.

**Definition 12 (Approximate reflection-symmetric or approximate translation-symmetric polygons)** *Let  $\delta_8$  and  $\delta_9$  be the permissible deviations from exact reflection-respective translation-symmetry. A polygon is approximate reflection-symmetric, if  $D_{RS} \leq \delta_8$ . A polygon is approximate translation-symmetric, if  $D_{TS} \leq \delta_9$ .*

Finally, we consider the right-angled polygon as a third type of typical regularity in man-made objects. In the special case of convex, right-angled polygons the shape is a rectangle. In general, the polygon can include concavities with the inner polygon angle  $270^\circ$ .

**Definition 13 (Right-angled polygon, right-angle deviation)** *A polygon is right-angled, if for every convergence of two line segments the inner polygon angle is constrained by  $d_{ra}(\gamma_i) = 0^\circ$ ,*

$$d_{ra}(\gamma_i) := \min\{|\gamma_i - 90^\circ|, |\gamma_i - 270^\circ|\} \quad (28)$$

*The right-angle deviation of a polygon with  $M$  convergences is defined by*

$$D_{RA}(\mathcal{G}) := \frac{1}{M * 360^\circ} * \sum_{i=1}^M d_{ra}(\gamma_i) \quad (29)$$

**Definition 14 (Approximate right-angled polygon)** *Let  $\delta_{10}$  be the permissible deviation from right-angled polygons. A polygon is approximate right-angled, if  $D_{RA} \leq \delta_{10}$ .*

Based on these definitions we introduce three quasi-invariants under projective transformation. For the imaging conditions in our experiments a threshold value  $\delta_8 = \delta_9 = \delta_{10} = 0.1$  proved as appropriate. The quasi-invariants will be used later on for extracting regular polygons from the image.

**Assumption 7 (Quasi-invariance of reflection-symmetry)** *The reflection-symmetry is a quasi-invariant, if a reflection-symmetric 3D polygon is approximate reflection-symmetric after projective transformation.*

**Assumption 8 (Quasi-invariance of translation-symmetry)** *The translation-symmetry is a quasi-invariant, if a translation-symmetric 3D polygon is approximate translation-symmetric after projective transformation.*

**Assumption 9 (Quasi-invariance of right-angles)** *The right-angle is a quasi-invariant, if a right-angled 3D polygon is approximate right-angled after projective transformation.*

The projective transformation of 3D object faces, which are reflection-symmetric, translation-symmetric or right-angled polygons, yields approximations of these specific polygons in the image. The features related to the geometric deviation of polygons from these specific shapes must be combined with features related to the geometric/photometric consensus for polygons. This gives a measure of conspicuity of specific polygons in an image.

**Definition 15 (Saliency of specific polygons)** *The saliency of a specific polygon is defined by*

$$S_{SP\_PG} := 1 - (\lambda_1 * D_{LE\_PG} + \lambda_2 * D_{CC\_PG} + \lambda_3 * D_{SP\_PG}) \quad (30)$$

The function symbol  $D_{SP\_PG}$  must be replaced by  $D_{RS}$ ,  $D_{TS}$ , or  $D_{CO}$ , depending on one is interested in approximate reflection-symmetric, translation-symmetric or right-angled polygons.

**Generic procedure for the extraction of specific polygons:**

1. From the whole set of combinations of three Hough peaks:
  - 1.1. Select just the combinations under the constraint that first and third Hough peak don't belong to the same cluster as the second peak.
  - 1.2. Determine for each combination a line segment by intersecting first and third line with the second one (specified by the Hough peaks respectively).
  - 1.3. Select the line segments, which are completely contained in the image, and are not isolated.
  - 1.4. Compute the line/edge orientation-deviation using function  $D_{LE}$ , and the convergence/corner junction-deviation using function  $D_{CC}$ , and select those line segments, for which both the LEOC and CCJC principles hold.
2. Compute a graph representing the neighborhood of line segments, i.e. create a knot for each intersection point and an arc for each line segment.
3. Compute the set of minimal, planar cycles in the graph, i.e. minimal numbers of knots and no arc in the graph is intersecting the cycles. This gives a candidate set of polygons representing faces of an object.
4. For each polygon:
  - 4.1. Compute the mean line/edge orientation-deviation using function  $D_{LE\_PG}$ .
  - 4.2. Compute the mean convergence/corner junction-deviation using function  $D_{CC\_PG}$ .
  - 4.3. Compute the deviation from a specific regularity using generic function  $D_{SP\_PG}$ .
  - 4.4. Compute the saliency value by combining the above results according to Eq. (30).
5. Bring the specific polygons into order according to decreasing saliency values.

This generic procedure has been applied successfully for localizing regular polygons, which originate from the surfaces of man-made objects. For example, the side and top face of the

computer monitor in Fig. 25 have been extracted. They were determined most saliently as approximate reflection-symmetric and approximate translation-symmetric octagons, respectively. As a second example, the boundary of the electronic board in Fig. 26 has been extracted. It was determined most saliently as approximate right-angled hexagon. Further examples are presented in the next section in the framework of extracting arrangements of polygons. E.g. the complete arrangement of polygons for the computer monitor will be determined by extracting and slightly adjusting the polygons of the side, top, and front faces under the consideration of certain assembly level constraints.

## 4 Consensus-based assembly level grouping

The extraction of regular polygons can be considered as a first step of the higher goal of localizing certain objects and describing their boundaries in more detail. Geometric/photometric consensus features have been combined with geometric regularity features for defining a saliency measure of specific polygons in the image. A salient polygon may arise from the boundary of a single object face or of a whole object silhouette. In general, it is assumed that the surface of a man-made 3D object can be subdivided in several faces and from these only a subset will be observed, e.g. in the case of a parallelepiped just 3 plane faces are observed from a non-degenerate view point. The projective transformation of that kind of object surface should yield an arrangement of several polygons. We introduce two assembly level grouping criteria, i.e. the vanishing point constraint and the convergence invariance constraint, which impose imperative restrictions on the shape of an arrangement of polygons. These constraints are directly correlated to the three-dimensional (regular) nature of the object surface. The principles are demonstrated for objects of roughly polyhedral shape, i.e. local protrusion, local deepening, or round corners are accepted.

### 4.1 Focusing image processing on polygonal windows

The consensus and regularity features, used so far, just take basic principles of image formation and qualitative aspects of the shape of man-made objects into account. Though only general assumptions are involved various experiments have shown that the extracted polygons are a useful basis for applying techniques of detailed object detection and boundary extraction. For example, Fig. 26 showed an electronic board, which has been extracted in a cluttered environment as an approximate right-angled hexagon. Subsequent image processing can focus on the hexagon image window for detecting specific electronic components on the board. As another example, Fig. 19 showed electrical objects and a set of extracted approximate parallelograms, in which the rough silhouette boundary of the dummy box was included (see Fig. 20). Subsequent image processing can focus on this parallelogram image window for extracting a detailed boundary description.

This section concentrates on detailed boundary extraction and in this context the extracted polygons serve a further purpose. We need to examine, which type of geometric shape a polygon surrounds in order to apply a relevant approach for detailed boundary extraction. The task of roughly characterizing the object shape (e.g. polyhedral or curvilinear) can be solved by taking a further constraints into account. According to the principles underlying the procedure of extracting polygons (in Section 3), it is reasonable to assume that there are polygons included, which approximately describe the silhouette of interesting objects. This has also been confirmed by the polygons extracted from the images in Fig. 20 and Fig. 26.

**Assumption 10 (Silhouette approximation by salient polygons)** *For the set of interesting objects, depicted in an image, there are salient polygons, which approximate the silhouettes with a necessary accuracy  $\delta_{11}$ .*

Based on this assumption it will be expected that a large part of the polygon closely

touches the object silhouette. Therefore, the gray value structure of the interior of a polygon, mainly belonging to the appearance of one object, can be taken into account in various approaches of object or shape classification. For example, in a histogram-based approach an object is represented by taking several views from it, and computing histograms of gray values, edge directions, corner properties, cooccurrence features, or further filter responses. In an offline phase a set of objects with relevant shapes is processed and the histograms stored in a database. In the online phase histograms are computed from the interior of a polygon, and matched with the database histograms. Based on the criterion of highest matching score the type of object shape must be concluded in order to apply the relevant approach of boundary extraction, e.g. extraction of arrangements of polygons or alternatively curvilinear shapes. For example, from the gray value structure in the quadrangle image window in Fig. 20 it has been concluded that the extraction of a detailed arrangement of polygons is reasonable. We have mentioned (just briefly) our approach of classification using histograms. It was inspired by the work of Swain and Ballard [37]. A detailed description is beyond the scope of this paper.

The Fig. 20 will serve to demonstrate the principles underlying a generic procedure for extracting arrangements of polygons. Though the dummy object is located in a cluttered scene, the silhouette quadrangle is acting as a window and the procedure of boundary extraction is hardly detracted from the environment or background of the object. According to that, we introduce the *windowed orientation-selective Hough transformation (WOHT)*, which just considers the image in a polygonal window. The definition is quite similar to that of OHT (see Definition 3), except that the votes are only collected from a subset  $P_S$  of coordinate tuples, taken from the interior of the extracted polygon and extended by a small band at the border. The WOHT contributes to overcome the problem of confusing profusion of Hough peaks. For example, we can apply the WOHT to the quadrangle window in Fig. 20 containing the approximate parallelepiped object. The boundary line configuration for three visible faces should consist of nine line segments to be organized in three sets of three approximate parallel lines respectively. In the Hough image of WOHT nine peaks must be organized in three stripes with three peaks in it respectively. This constraint has to be considered in an approach for searching the relevant Hough peaks. However, a configuration like this is just a necessary characteristic but not a sufficient one for constructing the correct object boundary. Further principles and invariants of projective transformation will be considered for the purpose of extracting the relevant arrangement of polygons.

A basic procedure for extracting configurations of Hough peaks has already been mentioned in Section 3.1. It extracts a certain number of Hough peaks and groups them by considering only the line parameter  $\phi$ . Related to the necessary characteristic the procedure must be repeated experimentally until at least three clusters occur, which consist of at least three Hough peaks respectively. Another alternative procedure more carefully directs the search. It is looking for the global maximum peak and thus determines the first relevant horizontal stripe. Within the stripe a certain number of other maximum peaks must be localized. Then the stripe is erased completely and in this modified Hough image the next global maximum is looked for. The new maximum defines the second relevant stripe, in which once again the specified number of other maximum peaks are detected. Repeating the procedure a certain number of times gives the final configuration of Hough

peaks. For demonstration, this procedure has been applied to the window in Fig. 20. A configuration of nine Hough peaks organized in three stripes of respective three peaks yields the set of image lines in Fig. 30. Though the necessary characteristic of the peak configuration holds, it is impossible to construct the complete object boundary because important boundary lines are missing. Fortunately, a configuration of 12 Hough peaks organized in four stripes (see Fig. 31) yields a more complete list of relevant boundary lines (see Fig. 32). The next two subsections take consensus principles and assembly level grouping criteria into account for evaluating or adjusting image lines for object boundary construction.

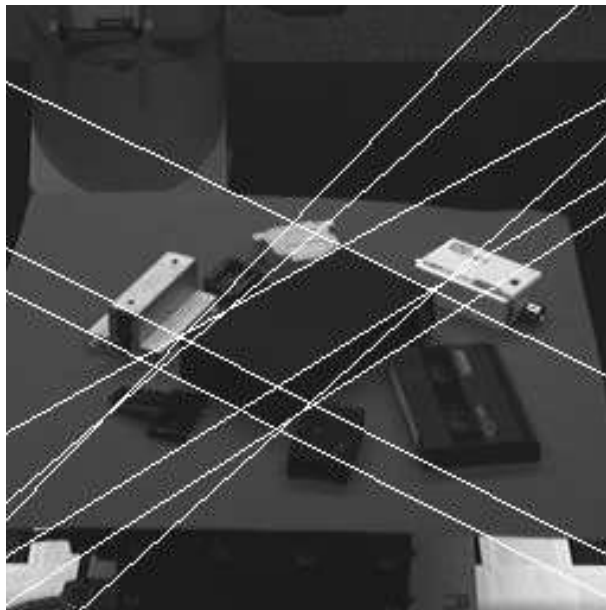


FIG. 30: Candidate set of nine boundary lines for the dummy box organized in three sets of three approximate parallel lines respectively. Result of applying the windowed OHT to the quadrangle image window in Fig. 20 and selecting nine Hough peaks organized in three stripes.

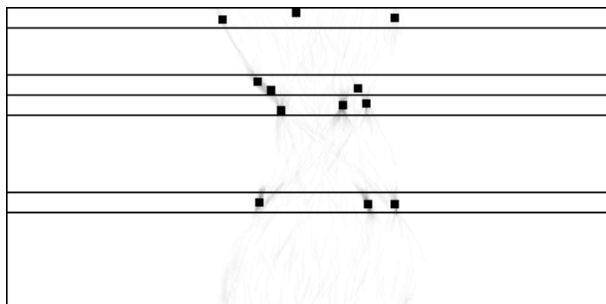


FIG. 31: Result of applying the windowed OHT to the quadrangle image window in Fig. 20 and selecting 12 Hough peaks organized in four stripes of respective three Hough peaks.



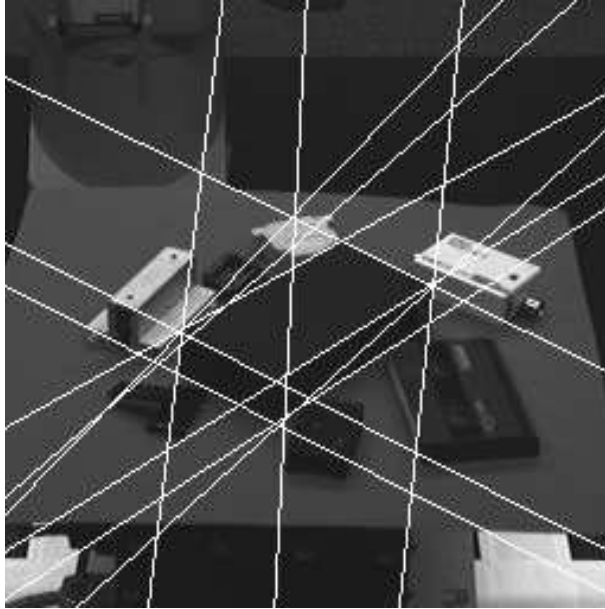


FIG. 32: Candidate set of 12 boundary lines for the dummy box specified by the 12 Hough peaks in Fig. 31. More relevant relevant boundary lines are included than in Fig. 30.

## 4.2 Vanishing point constraint of parallel boundary lines

The projective transformation of parallel boundary lines generates approximate parallel image lines with the specific constraint that they meet in one vanishing point  $p_v$ . This vanishing point constraint imposes certain qualitative constraints on the courses of Hough peaks within a horizontal stripe (specifying approximate parallel lines). Figure 33 shows a projected parallelepiped and two vanishing points  $p_{v1}$  and  $p_{v2}$  for two sets  $\{\mathcal{L}_{11}, \mathcal{L}_{12}, \mathcal{L}_{13}\}$  and  $\{\mathcal{L}_{21}, \mathcal{L}_{22}, \mathcal{L}_{23}\}$  of three approximate parallel line segments respectively. Let  $(r_{ij}, \phi_{ij})$  be the polar form parameters of the lines respectively. We realize for the monotonously increasing distance parameter  $r_{11} < r_{12} < r_{13}$  of the first set of lines a monotonously *increasing* angle parameter  $\phi_{11} < \phi_{12} < \phi_{13}$ , and for the monotonously increasing distance parameter  $r_{21} < r_{22} < r_{23}$  of the second set of lines a monotonously *decreasing* angle parameter  $\phi_{21} > \phi_{22} > \phi_{23}$ .

This specific observation is generalized to the following geometric regularity constraint.

**Assumption 11 (Vanishing point constraint)** *Let  $\{\mathcal{L}_1, \dots, \mathcal{L}_H\}$  be a set of approximate parallel line segments in the image, which originate from projective transformation of parallel line segments of the 3D object boundary. The extended lines related to the image line segments meet at a common vanishing point  $p_v$  and can be ordered according to the strong monotony  $r_1 < \dots < r_i < \dots < r_H$  of the distance parameter. For this arrangement there is a weak monotony of the angle parameter,*

$$\phi_1 \geq \dots \phi_i \geq \dots \geq \phi_H \quad \text{or} \quad \phi_1 \leq \dots \phi_i \leq \dots \leq \phi_H \quad (31)$$

We have to be careful with approximate vertical lines, whose angle parameter  $\phi$  is near to  $0^\circ$  or near to  $180^\circ$ . In a cluster of Hough peaks with that characterization all lines with  $\phi$  near to  $0^\circ$  will be redefined by:  $\hat{r} := -r$ , and  $\hat{\phi} := \phi + 180^\circ$ . This is permitted because

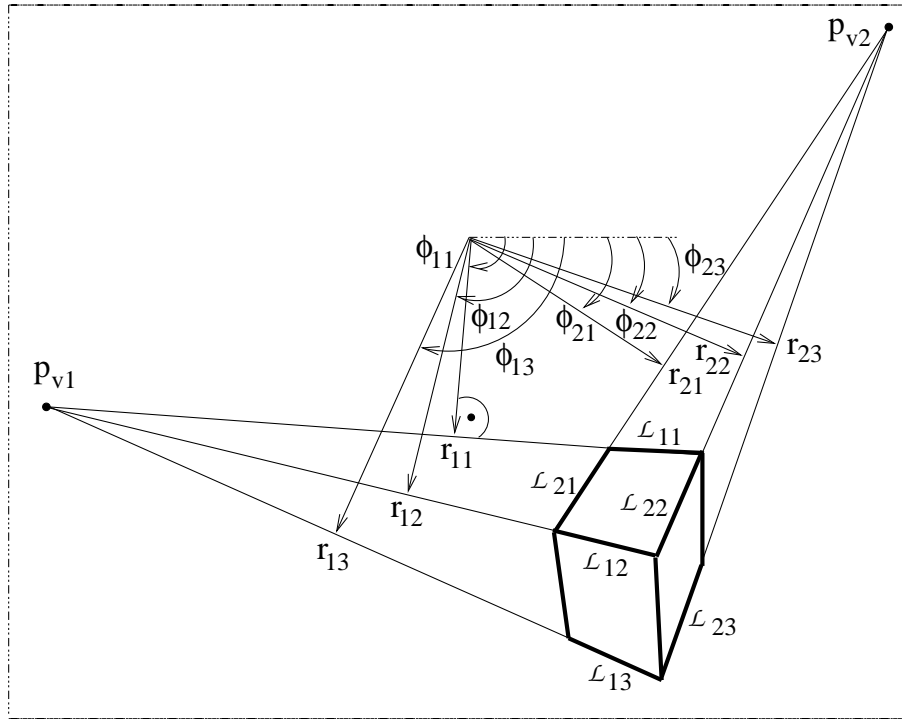


FIG. 33: Projected parallelepiped and two vanishing points  $p_{v1}$  and  $p_{v2}$ . Monotonously increasing angle parameter  $\phi_{11} < \phi_{12} < \phi_{13}$ , and monotonously decreasing angle parameter  $\phi_{21} > \phi_{22} > \phi_{23}$  for two sets of three approximate parallel line segments respectively.

the equations  $L(p, (r, \phi)) = 0$  and  $L(p, (-r, \phi + 180^\circ)) = 0$  of Eq. (1) define the same line (which is easily proven). Under this consideration the Assumption 11 must hold for any set of approximate parallel lines meeting at a common point. Consequently, the *course of Hough peaks* in a horizontal stripe must *increase* or *decrease* weak monotonously.

For demonstration, this constraint is examined in the Hough image of clustered peaks in Figure 31. Assumption 11 holds for the third and fourth stripe but not for the first and second stripe. Actually, the Hough peaks in the first stripe specify lines, which are candidates for the short boundary lines of the object (approximate vertical lines in Fig. 32). The problem arises for the middle line due to small gray value contrast between neighboring faces. The Hough peaks of the second stripe originate from neighboring objects at the border of the quadrangle image window.

The vanishing point constraint is useful for slightly modifying the parameters  $r$  and  $\phi$  of extracted image lines. A simple procedure is applied, which assumes that in a set of approximate parallel lines at least two lines are reliable and need not be adjusted. Candidates for this pair of *seed lines* are outer silhouette lines, which can be extracted robustly in case of high contrast between the gray values of object and background. Otherwise, two inner boundary lines of the silhouette could be seed lines as well, e.g. boundary lines of object faces in case of high gray value contrast due to lighting conditions or different face colors. The reliability of a line is computed on the basis of line/edge orientation-deviation in Definition 4. However, thus far the lines in Fig. 32 are not restricted to the relevant line segment of the object border. Therefore, we specify for each candidate line

(e.g. in Fig. 32) a virtual line segment, which is of the same orientation respectively. For the virtual segments a unique length is specified, which is assumed to be a lower bound of the lengths of all relevant boundary line segments in the image. Each virtual line segment will be moved in discrete steps along the affiliated candidate line from border to border of the polygonal window. Step by step the orientation-consensus is evaluated by applying Eq. (5). The minimum is taken as the reliability value of the line.

The most reliable two lines are selected as seed lines and their point of intersection computed, which is taken as the vanishing point. Next, the other approximate parallel lines (which are less reliable) are redefined such that they intersect at the vanishing point. The redefined lines are slightly rotated around the vanishing point in order to optimize the reliability value under the constraint that the weak monotony constraint must hold in the course of Hough peaks of all approximate parallel lines. Exception handling is necessary, if the two seed lines are exact parallel, because there is no finite vanishing point. In this case the unique orientation from the seed lines is adopted for the less reliable lines and a slight translation is carried out (if necessary) to optimize their reliability values. In order to take the geometric/photometric consensus into account the seed lines and/or the redefined lines are only accepted if the line/edge orientation-consensus holds (see Assumption 2), otherwise they are discarded.

For example, this procedure can be applied to the set of three approximate vertical lines in Fig. 32, represented by three non-monotonous Hough peaks in the first stripe of Fig. 31. As a result, the two outer lines have been determined as seed lines, and the inner line is slightly rotated to meet the vanishing point constraint. The next subsection introduces a further constraint inherent in the projection of polyhedral objects, which will be applied in combination with the vanishing point constraint later on.

### 4.3 Convergence invariance of meeting boundary lines

In man-made objects the most prominent type of junction is a convergence of three lines ( $(\#3)$ -junction) respectively, e.g. a parallelepiped includes eight  $(\#3)$ -junctions. By means of projective transformation some parts of an opaque object boundary will be occluded, which makes certain junctions just partly visible or even invisible. For example, under general view conditions we realize in the image of a parallelepiped four  $(\#3)$ -junctions and three  $(\#2)$ -junctions (see Figure 33). I.e. in four junctions all three converging lines are visible, in three junctions only two lines are visible respectively, and one junction is completely occluded. A thorough analysis of visibility aspects of polyhedral objects was presented by D. Waltz for the purpose of interpreting line drawings (see [38], pp. 249-281). We introduce a geometric invariance constraint related to the  $(\#3)$ -junctions, for which three converging lines are visible in the image.

**Assumption 12 (Convergence invariance constraint)** *Let a 3D convergence point be defined by the intersection of three meeting border lines of an approximate polyhedral 3D object. The projective transformation of the 3D convergence point should yield just one 2D convergence point in the image.*

For demonstration we select from the image in Fig. 32 a subset of three boundary lines and compute the intersection points as shown in Fig. 34. Obviously, Assumption 12 does

not yet hold because just one common intersection point is expected instead of three. The reason is that line extraction via Hough transformation is more or less inaccurate (like any other approach to line extraction). Actually, correctness and accuracy of lines can only be evaluated with regard to the higher goal of extracting the whole object boundary. The previously introduced vanishing point constraint has been a first opportunity of including high level goals to line extraction, and the constraint of convergence invariance is a second one.

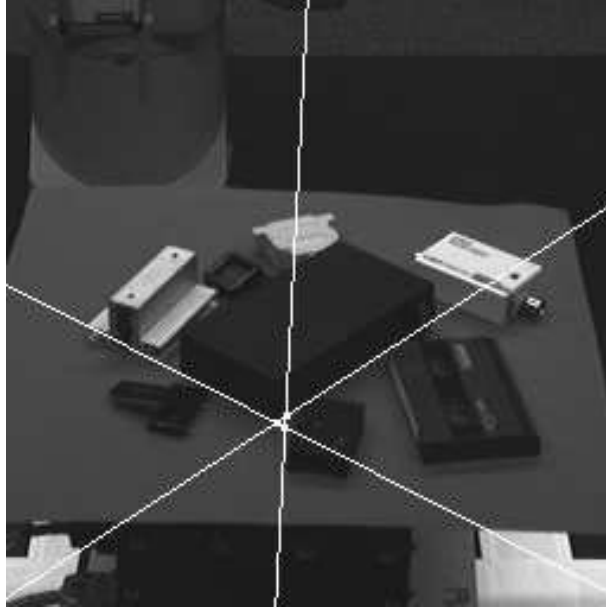


FIG. 34: Subset of three boundary lines taken from Fig. 32 and three different intersection points. One unique intersection point is requested to fulfill the convergence invariance constraint.

In order to make Assumption 12 valid we apply a simple procedure, which only adjusts the position parameter  $r$  of image lines. The idea is to select from a (#3)-junction the most reliable two lines (using the procedure mentioned above), compute the intersection point, and translate the third line into this point. The approach proved to be reasonable because of our frequent observation, that two lines of a (#3)-junction are acceptable accurate and sometimes just the third line is deviating to a larger extent. For example, the most reliable two lines in Fig. 34 are the slanted ones, and therefore the intersection point is computed and the approximate vertical line is parallel translated into this point. More sophisticated procedures are conceivable, which flexibly fine-tune the parameters of several relevant lines in combination (not treated in this work).

#### 4.4 Boundary extraction for approximate polyhedra

The geometric/photometric consensus constraints and the geometric grouping criteria at the primitive, the structural, and the assembly level can be combined in a generic procedure for extracting the arrangement of polygons for a polyhedral object boundary. A precondition for the success of this procedure is that all relevant line segments, which are included in the arrangement of polygons can be detected as peaks in the Hough image.

**Assumption 13 (High gray value contrast between object faces)** *All transitions between neighboring faces of a polyhedral object are characterized by high gray value contrast of at least  $\delta_{12}$ .*

**Generic procedure for the extraction of arrangements of polygons:**

1. Apply the windowed OHT in a polygonal image window, detect a certain number of Hough peaks, and consider that they must be organized in stripes.
2. For each stripe of Hough peaks, examine the vanishing point constraint, and if it does not hold, then apply the procedure mentioned previously.
3. Compute intersection points for those pairs of image lines, which are specified by pairs of Hough peaks located in different stripes.
4. Determine all groups of three intersection points located in a small neighborhood respectively, and for each group examine the constraint of convergence invariance and if it does not hold then apply the procedure mentioned previously.
5. Based on the redefined lines determine a certain number of most salient polygons (see Definition 15) by applying a procedure similar to the one presented in Section 3.
6. Group the polygons into arrangements, compute an *assembly value* for each arrangement, and based on that select the most relevant arrangement of polygons.

The pre-specified number of peaks to be extracted in the first step of the procedure must be high enough such that all relevant peaks are included. Another critical parameter is involved in the fifth step, i.e. extracting a certain number of most salient polygons. We must be careful that all relevant polygons are included, which are needed for constructing the complete arrangement of polygons of the object boundary. The final step of the procedure will be implemented dependent on specific requirements and applications. For example, the grouping of polygons can be restricted to arrangements, which consist of connected, non-overlapping polygons of a certain number, e.g. arrangements of 3 polygons for describing three visible faces of an approximate parallelepiped. The *assembly value* of an arrangement of polygons can be defined as the *mean saliency value* of all included polygons.

With this specific realization the generic procedure has been applied to the quadrangle image window in Fig. 20. The extracted boundary for the object of approximate parallelepiped shape is shown in Fig. 35. Further, the procedure has been applied to more general octagons, e.g. a loudspeaker, whose surface consists of rectangles and trapezoids (see Fig. 36). The procedure also succeeds for more complicated shapes like the computer monitor, which has already been treated in the previous section (see Fig. 25). The extracted boundary in Fig. 37 demonstrates the usefulness of the convergence invariance

constraint. There are four (#3)-junctions, which have a unique convergence point respectively (as opposed to the partial boundary in Fig. 25).

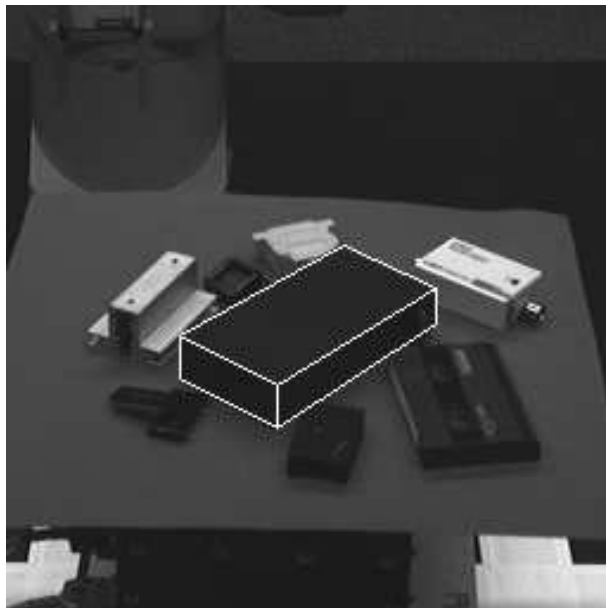


FIG. 35: Dummy box with approximate right-angled parallelepiped shape in a complex environment. Arrangement of polygons describing the visible boundary of the dummy box.



FIG. 36: Loudspeaker with approximate octagonal shape in a complex environment. Arrangement of polygons describing the visible boundary.

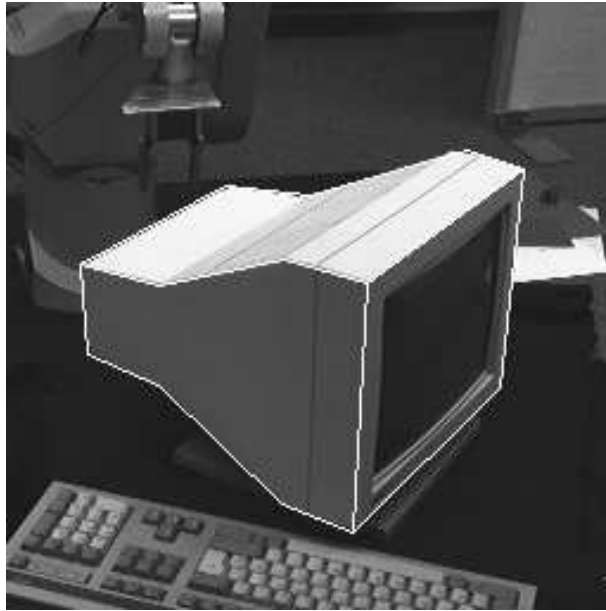


FIG. 37: Computer monitor with approximate polyhedral shape including non-convexities. Arrangement of polygons describing the visible boundary.

Though the procedure for boundary extraction reveals impressive results for more or less complicated objects, however it can fail in simple situations. This is due to the critical assumption that all line segments of the boundary must be detected explicitly as peaks in the Hough image. For images of objects with nearly homogeneous surface color, like the dummy box in Fig. 35, the contrast between faces is just based on lighting conditions, which is an unreasonable basis for boundary extraction. On the other hand, for objects with texture or inscription on the surface spurious gray value edges exist, which are as distinctive as certain relevant edges at the border of the object silhouette. However, all linear edge sequences produce a Hough peak respectively. As a consequence, perhaps a large number of Hough peak must be extracted such that all relevant boundary lines are included.

#### 4.5 Sophisticated geometric reasoning for boundary extraction

This section presents a modified procedure for boundary extraction, which applies a sophisticated strategy of geometric reasoning. It is more general in the sense that the critical Assumption 13, involved in the procedure presented above, is weakened. However, boundary extraction is restricted to objects of approximate parallelepiped shape and therefore the procedure is more specific concerning the object shape. The usability of the procedure is based on the following assumption.

**Assumption 14 (Parallelepiped approximation)** *The reasonable type of shape approximation for the object in a quadrangle image window is the parallelepiped.*

## Extraction of arrangements of polygons for parallelepipeds:

1. Determine a quadrangle image window, which contains an object of approximate parallelepiped shape.
2. Determine just the boundary of the object silhouette, which is assumed to be the most salient hexagon.
3. Propagate the silhouette lines (outer boundary lines) to the interior of the silhouette to extract the inner lines. Apply the geometric/photometric consensus criteria and the assembly level grouping criteria to extract the most relevant arrangement of polygons.

Fig. 38 shows the quadrangle image window containing the relevant target object, i.e. a transceiver box of approximate parallelepiped shape. The boundary line segments of the parallelepiped silhouette must form a hexagon (see Fig. 39). A saliency measure is defined for hexagons, which takes into account the structural level grouping criterion of reflection-symmetry and the aspect that the hexagon must touch a large part of the quadrangle contour. This reveals the boundary line segments in Fig. 39, which are organized as three pairs of two approximate parallel line segments respectively. Additionally, three inner line segments of the silhouette are needed to build the arrangement of polygons for the boundary of the parallelepiped. The vanishing point constraint is taken into account to propagate the approximate parallelism of outer lines to the interior of the silhouette. Furthermore, the convergence invariance constrains inner lines to go through the convergence points of the silhouette boundary lines and additionally to intersect in the interior of the silhouette at just one unique point. The final arrangement of polygons must consist of just four (#3)-junctions and three (#2)-junctions. The combined use of the assembly level criteria guarantees that only two configurations of three inner lines are possible (one configuration is shown in Fig. 40). The relevant set of three inner line segments is determined based on the best geometric/photometric consensus. Fig. 41 shows the final boundary line configuration for the transceiver box.



FIG. 38: Transceiver box with approximate right-angled parallelepiped shape. The black quadrangle surrounding the object indicates the image window for detailed processing.





FIG. 39: Extracted regular hexagon, which describes the approximate silhouette of the transceiver box.

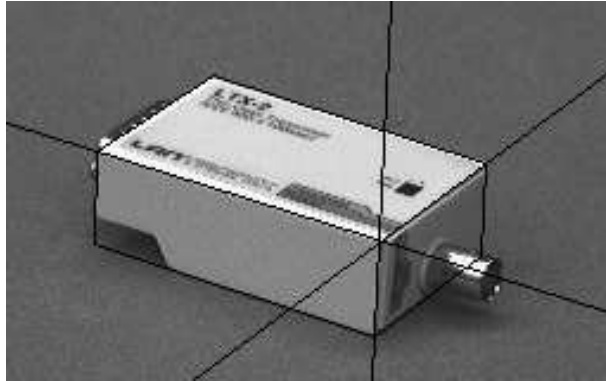


FIG. 40: Relevant set of three inner lines of the silhouette of the transceiver box. They have been determined by propagation from outer lines using assembly level grouping criteria and the geometric/photometric consensus.

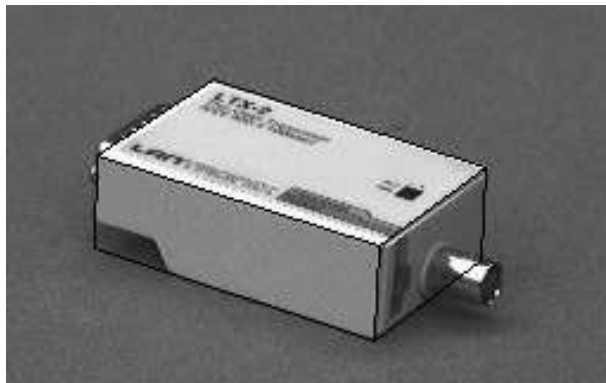


FIG. 41: Transceiver box with final polygon arrangement for the parallelepiped boundary description.

Further examples of relevant object boundaries are given below (see Fig. 42 and Fig. 43). They have been extracted from usual images of electrical scrap using the procedure just introduced.

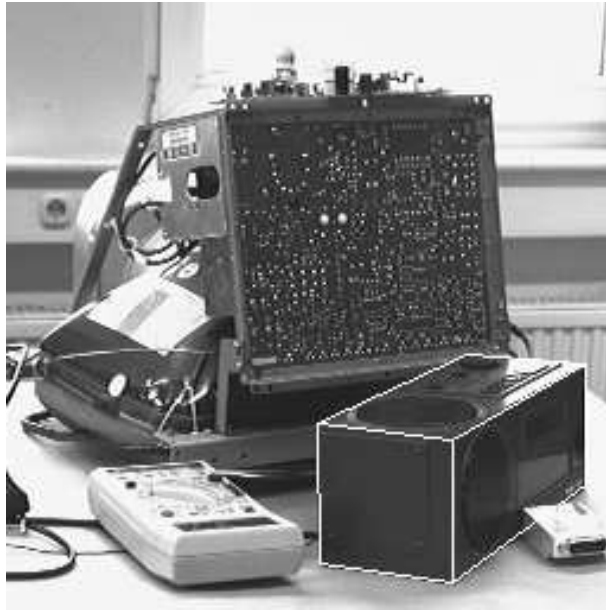


FIG. 42: Radio with approximate right-angled parallelepiped shape and extracted arrangement of polygons of the boundary.

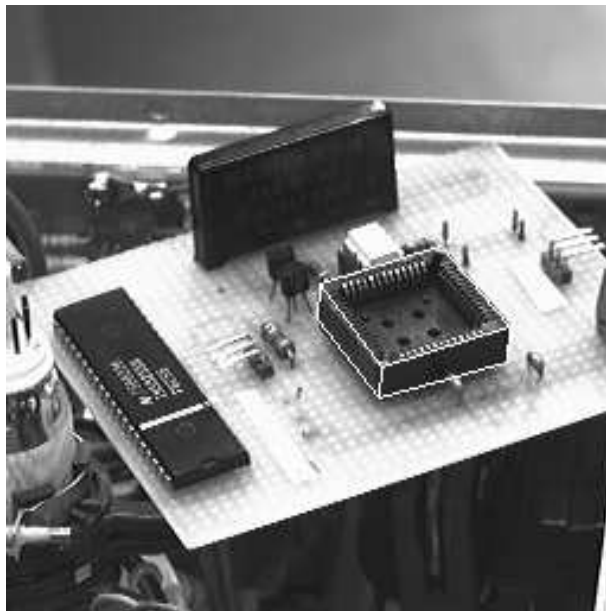


FIG. 43: Chip-carrier with approximate right-angled parallelepiped shape and extracted arrangement of polygons of the boundary.

## 5 Discussion and conclusion

The system for boundary extraction is organized in several generic procedures, for which the relevant definitions, assumptions, and realizations have been presented. In general, it works successful for man-made objects of approximate polyhedral shape. Interestingly, the various assumptions can be organized into three groups by considering the level of generality.

- The first group introduces general geometric/photometric consensus principles for polyhedral objects.
- The second group considers shape regularity features, which are quasi-invariants for a subset of viewpoints.
- The third group incorporates more specific assumptions concerning object appearance and shape.

Therefore, the three groups of assumptions are stratified according to decreasing generality, which imposes a certain level of speciality on the procedures.

The first group consists of the Assumptions 1, 2, 3, 5, and 6. These are based on functions for evaluating the geometric/photometric consensus related to a line point, a line segment, a line convergence, a quadrangle or an arbitrary polygon. Threshold parameters  $\delta_1$ ,  $\delta_2$ ,  $\delta_3$ ,  $\delta_4$ ,  $\delta_6$ ,  $\delta_7$  are involved for specifying the necessary geometric/photometric consensus. These criteria can be used for accepting just the relevant line structures in order to increase the efficiency of subsequent procedures for boundary extraction. According to our experience the parameters can be determined in a training phase prior to the actual application phase. They depend mainly on the characteristics of image processing techniques involved and of the camera objectives used. For example, we must clarify in advance the accuracy of the orientation of gray value edges and the accuracy of the localization of gray value corners, and related to the process of image formation, we are interested in the field of sharpness and the distortion effects on straight lines. Based on these measurements the threshold parameters are specified. In principle, the assumptions of this first group are valid for arbitrary polyhedral objects, from which images are taken with usual camera objectives.

The second group consists of the Assumptions 4, 7, 8, 9, 11, 12. They impose constraints on the projective transformation of geometric features of 3D object shapes. To consider the regularity aspect of man-made objects a set of collated regularity features is used, like parallel lines, right-angled lines, reflection-symmetric polylines, or translation-symmetric polylines. The object shapes are detected in the image as salient polygons or arrangements of polygons. Several saliency measures have been defined on the basis of geometric/photometric consensus features and the collated regularity features (just mentioned). Therefore, it is essential that the regularity features are invariant or at least quasi-invariant under projective transformation. The degree of deviation from exact invariance depends on the spectrum of permissible camera positions relative to the scene objects. Threshold parameters  $\delta_5$ ,  $\delta_8$ ,  $\delta_9$ ,  $\delta_{10}$  are involved in the assumptions for describing the permissible degrees of deviation from exact invariance. For example, if we would like to locate the right-angled silhouette of a flat object (e.g. an electronic board), then

the camera must be oriented approximately perpendicular to that object, and this will be considered in the parameter  $\delta_{10}$  (see Assumption 9). In principle, the assumptions of this second group are valid for arbitrary polyhedral objects but restrict the possible view conditions.

The basic assumption that the scene consists of approximate polyhedral objects usually is too general for providing one and only one generic procedure for boundary extraction. Therefore, a third group of constraints is introduced consisting of the Assumptions 10, 13, 14. They impose constraints on the gray value appearance and the shape of the depicted objects. We must examine whether an extracted polygon is an approximate representation of the object silhouette, or examine whether the transitions between object faces have high gray value contrast, or examine whether the shape of an object in a quadrangle image window is an approximate parallelepiped. Threshold parameters  $\delta_{11}$  and  $\delta_{12}$  are involved for quantifying these constraints. Though the assumptions of this third group are more specific than those of the other two groups, they are somewhat general.

Altogether, our system succeeds in locating and extracting the boundary line configurations for approximate polyhedral objects in cluttered scenes. Following Occam's minimalistic philosophy, the system makes use of fundamental principles underlying the process of image formation, and makes use of general regularity constraints of man-made objects. Based on that the role of specific object models is reduced. This aspect is useful in many realistic applications, for which it is costly or even impossible to acquire specific object models. For example, in the application area of robotic manipulation of electrical scrap (or car scrap, etc.), it is inconceivable and anyway unnecessary to explicitly model all possible objects in detail. For robotic manipulation of the objects approximate polyhedral descriptions are sufficient, which can be extracted on the basis of general assumptions. The novelty of our methodology is that we maximally apply general principles and minimally use object-specific knowledge for extracting the necessary information from the image to solve a certain task.

Future work should discover more consensus features between geometry and photometry of image formation, more regularity features of objects, and more quasi-invariance constraints. The combination of consensus, regularity, and quasi-invariance features must be treated thoroughly, e.g. solving the problem of combined constraint satisfaction. An extension of the methodology beyond man-made objects, e.g. natural objects like faces, is desirable.

**Acknowledgement:** The discussions with G. Sommer are greatly appreciated.

# List of symbols

$i$	Index
$x_1, x_2, y_1, y_2, y_3$	Image and scene coordinates
$p, p_i, p_a, p_b, \dots$	Image coordinate tuples
$P, P_S$	Set or subset of image coordinate tuples
$r$	Distance of a line from image center
$\phi$	Orientation of a line
$q$	Parameter tuple of a line
$Q$	Set of parameter tuples of lines
$L$	Polar line representation
$I_w, I_h, I_d$	Width, height, diagonal of an image
$B, O$	Binary image, and image of edge orientations
$H, K, M, N$	Upper bounds of indices
$f$	Effective focal length
$\delta_i$	Threshold parameters for various purposes
$\lambda_i$	Weighting factors
$\Delta a$	Tolerance band for edge orientation
$\mathcal{L}$	Sequence of discrete points describing a line segment
$(\#M)$ -junction	Junction with M converging lines
$\alpha_i, \beta_i, \gamma_i, \phi_i$	Angles for lines and convergences
$s_i$	Lengths of line segments
$\mathcal{A}, \mathcal{B}, \mathcal{F}, \mathcal{G}$	Sequences of angles
$\mathcal{S}$	Sequence of lengths
$D_O$	Absolute distance between angles modulo 180
$D_{LE}$	Orientation-deviation related to a line segment
$D_{CC}$	Junction-deviation related to a line convergence
$D_{JP}$	Euclidean distance between positions
$D_{JO}$	Deviation between two sequences of angles
$D_{LE}$	Orientation-deviation related to a quadrangle
$D_{CC}$	Junction-deviation related to a quadrangle
$D_{RC}$	Rectangle-deviation of a quadrangle
$D_{PA}$	Parallelogram-deviation of a quadrangle
$D_{SQ}$	Square-deviation of a quadrangle
$D_{RH}$	Rhombus-deviation of a quadrangle
$D_{TR}$	Trapezoid-deviation of a quadrangle
$D_{SP\_QD}$	Generic measure for deviation from specific quadrangle
$S_{SP\_QD}$	Saliency measure of a specific quadrangle
$D_{RS}, d_{rs}$	Deviation from reflection-symmetry
$D_{TS}, d_{ts}$	Deviation from translation-symmetry
$D_{RA}, d_{ra}$	Deviation from a right-angled polygon
$V_{SL}$	Normalized length variance of line segments

# References

- [1] Y. Aloimonos. Active vision revisited. In Y. Aloimonos, editor, *Active Perception*, pages 1–18. Lawrence Erlbaum Associates Publishers, New Jersey, 1993.
- [2] K. Shimoga. Robot grasp synthesis algorithms: A survey. *The International Journal of Robotics Research*, 15:230–266, 1996.
- [3] O. Faugeras. Stratification of three-dimensional vision: Projective, affine and metric representations. *Journal of the Optical Society of America*, 12:465–484, 1985.
- [4] V. Leavers. Survey: Which Hough transform ? *Computer Vision, Graphics, and Image Processing – Image Understanding*, 58:250–264, 1993.
- [5] New York, June 14. *IEEE Workshop on Qualitative Vision*. IEEE Computer Society Press, Los Alamitos, USA, 1993.
- [6] R. Bajcsy and M. Campos. Active and exploratory perception. *Computer Vision and Image Understanding*, 56:31–40, 1992.
- [7] G. Sommer. Algebraic aspects of designing behavior based systems. In G. Sommer and J. Koenderink, editors, *Algebraic Frames for the Perception-Action Cycle*, volume 1315 of *Lecture Notes in Computer Science*, pages 1–28, Berlin, 1997. Springer Verlag.
- [8] R. Haralick and L. Shapiro. *Computer and Robot Vision*. Addison-Wesley Publishing Company, Massachusetts, 1993. volume I.
- [9] D. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, Boston, USA, 1985.
- [10] J. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. The MIT Press, Cambridge, Massachusetts, 1992.
- [11] T. Binford and T. Levitt. Quasi-invariants: Theory and exploitation. In *Image Understanding Workshop*, pages 819–829, 1993.
- [12] P. Gros, O. Bournez, and E. Boyer. Using local planar geometric invariants to match and model images of line segments. *Computer Vision and Image Understanding*, 69:135–155, 1998.
- [13] O. Faugeras. *Three-Dimensional Computer Vision*. The MIT Press, Cambridge, Massachusetts, 1993.
- [14] R. Jain and T. Binford. Ignorance, myopia, and naiveté in computer vision systems. *Computer Vision and Image Understanding*, 53:112–117, 1991.
- [15] D. Crevier and R. Lepage. Knowledge-based image understanding systems: A survey. *Computer Vision and Image Understanding*, 67:161–185, 1997.
- [16] D. Marr. *Vision – A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman and Company, New York, USA, 1982.
- [17] D. Ballard and C. Brown. Principles of animate vision. *Computer Vision and Image Understanding*, 56:3–21, 1992.

- [18] S. Sarkar and K. Boyer. Perceptual organization in computer vision: A review and a proposal for a classificatory structure. *IEEE Transactions on Systems, Man, and Cybernetics*, 23:382–399, 1993.
- [19] S. Sarkar and K. Boyer. A computational structure for preattentive perceptual organization - Graphical enumeration and voting methods. *IEEE Transactions on Systems, Man, and Cybernetics*, 24:246–267, 1994.
- [20] P. Palmer, J. Kittler, and M. Petrou. An optimizing line finder using a Hough transform algorithm. *Computer Vision and Image Understanding*, 67:1–23, 1997.
- [21] J. Princen, J. Illingworth, and J. Kittler. A hierarchical approach to line extraction based on the Hough transform. *Computer Vision, Graphics, and Image Processing*, 52:57–77, 1990.
- [22] F. Wahl and H. Biland. Decomposing of polyhedral scenes in Hough space. In *International Conference on Pattern Recognition*, pages 78–84. Paris, 1989.
- [23] D. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13:111–122, 1981.
- [24] M. Yang, J.-S. Lee, C.-C. Lien, and C.-L. Huang. Hough transform modified by line connectivity and line thickness. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:905–910, 1997.
- [25] G. Foresti, V. Murino, C. Regazzoni, and G. Vernazza. Grouping of rectilinear segments by the labeled Hough transform. *Computer Vision and Image Understanding*, 58:22–42, 1994.
- [26] K. Cho and P. Meer. Image segmentation from consensus information. *Computer Vision and Image Understanding*, 68:72–89, 1997.
- [27] A. Amir and M. Lindenbaum. A generic grouping algorithm and its quantitative analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:168–185, 1998.
- [28] A. Castano and S. Hutchinson. A probabilistic approach to perceptual grouping. *Computer Vision and Image Understanding*, 64:399–419, 1996.
- [29] A. Ylä-Jääski and F. Ade. Grouping symmetric structures for object segmentation and description. *Computer Vision and Image Understanding*, 63:399–417, 1996.
- [30] A. Zisserman, D. Forsyth, and J. Mundy. 3D object recognition using invariance. *Artificial Intelligence*, 78:239–288, 1995.
- [31] G. Granlund and H. Knutsson, editors. *Signal Processing for Computer Vision*. Kluwer Academic Publishers, 1995.
- [32] K. Rohr. Localization properties of direct corner detectors. *Journal of Mathematical Imaging and Vision*, 4:139–150, 1994.
- [33] S. Smith and J. Brady. SUSAN: A new approach to low level image processing. *International Journal of Computer Vision*, 23:45–78, 1997.
- [34] E. Simoncelli and H. Farid. Steerable wedge filters for local orientation analysis. *IEEE Transactions on Image Processing*, 5:1377–1382, 1996.

- [35] P. Havaldar, G. Medioni, and F. Stein. Perceptual grouping for generic recognition. *International Journal of Computer Vision*, 20:59–80, 1996.
- [36] R. Schalkoff. *Pattern Recognition - Statistical, Structural, and Neural Approaches*. John Wiley and Sons, New York, 1992.
- [37] M. Swain and D. Ballard. Color indexing. *International Journal of Computer Vision*, 7:11–32, 1991.
- [38] P. Winston. *Artificial Intelligence*. Addison-Wesley Publishing Company, Reading Massachusetts, 1992.