Projective Invariance and Orientation Consensus for Extracting Boundary Configurations

Josef Pauli

Christian–Albrechts–Universität, Institut für Informatik
Preusserstrasse 1-9, 24105 Kiel, Germany

Abstract. This work focusses on boundary detection of target objects by extracting configurations of straight lines. Hough transformation is used as basic voting technique for extracting the image lines. To find the appropriate peaks in the bin array of the Hough transform and thus estimating the line parameters a three-step procedure is involved. First, the Hough transform is constrained appropriately by the local orientations of grey level edges. Second, the global histogram of edge orientations is used to more or less emphasize certain parts in the Hough image. With these two modifications of the standard Hough transform we incorporate principles of orientation consensus and thus sharpen the peaks in the parameter space. Third, parallelism as a quasi-invariant of perspective projection is used to apply a sophisticated search strategy. Throughout the approach no specific object models but only basic principles of the imaging process are involved in extracting the boundary lines.

1 Introduction

In many applications of computer vision it is sufficient to describe the scene objects by polyhedral approximations, e.g. in vision based robot grasping [2]. For reconstructing the line frame it is indispensible to detect features like edges, corners or line segments in the image arising from the surface of an object. For example, the authors in [3] extract relevant image lines by using model lines and fitting the image gradients to the model gradients. The model lines are computed based on the two-dimensional projection of discontinuities between the faces of a 3D polyhedral model. The next task is to associate those line segments in the image that arise from an individual object. The authors in [7] extract the outer border of an object (silhouette) by using typical techniques of edge detection, contour following and polygonal approximation. The key idea is then to make use of perspective invariants for deducing the interior boundary lines of the object. For example, under affine imaging conditions the parallelism of 3D lines of a polyhedra must also hold between the projected lines in the image. Accordingly, certain lines of the outer border of a rectangular solid appear with the same orientation in the interior of the solid silhouette.

We implemented a system for computing from an image the two-dimensional *boundary line configuration* which arises from the three-dimensional border lines of a scene object. Principles of orientation consensus and projective invariance are applied but unlike the cited papers no specific object models are involved.

A modified Hough transformation is used as basic voting technique for determining lines. In contrast to usual line regression the Hough transformation [4] can lead to a *robust estimation* of the line parameters which are not distorted by outliers in the edge image. Each line corresponds to a peak in the Hough space of line parameters. The most important problem is to find especially those peaks which describe the boundary lines of an individual object. Our paper presents an approach for finding these sets of appropriate peaks. For the purpose of demonstration a rectangular solid is used, but the system is applicable more generally. The procedure for sharpening the peaks is useful for any polyhedral object. The strategy for searching the peaks works for any parallelepiped having two parallelograms as top and bottom faces and four parallelograms as side faces. It is easily conceivable to implement further strategies in order to extract the boundary of other object shapes.

2 Edge orientation based Hough transformation

First step: *Computing local orientations and gradient magnitudes*
For purpose of illustration we take a subimage where a target object is located in (see Figure 1, left). Let X be the set of coordinate tupels $p = (x1, x2)$ of this subimage I_g. We regard every point in the subimage as edge point and determine the *local orientation* $O_a(p)$. A standard approach is used by combining four differently oriented 2D–Gaborfunctions which respond sensitively only to certain directions of the edges. We take the energy of the complex response to measure the strength of edges along a certain direction [1]. The choice of orientations is $0°$, $45°$, $90°$, and $135°$, which means that the filters respond most sensitive to edges whose orientations are colinear to the given angles. Let the energy of the complex response of four filters be $E_1(p), E_2(p), E_3(p), E_4(p)$ at each point. The energies are multiplied with the cosine of the doubled angle and added up, and this procedure is repeated with the sine of the doubled angles (symbols C and S for cosine and sine in the formula below). From the two results we compute the arcus tangens (symbol At in the formula below) taking the quadrant of the coordinate system into account and do simple exception handling at singularities. For determining the orientation of the edge we have to multiply the factor -0.5 and if the resulting angle is negative then we add π. The orientation is in radian values between between 0 and π where the turning direction is clockwise.[1] Finally, by normalization we transform the orientation into a discrete angle degree of the integer set $\{0, \cdots, 179\}$.

$$E_1(p)\mathrm{C}(0°)+E_2(p)\mathrm{C}(90°)+E_3(p)\mathrm{C}(180°)+E_4(p)\mathrm{C}(270°) = E_1(p)-E_3(p) \quad (1)$$

$$E_1(p)\mathrm{S}(0°)+E_2(p)\mathrm{S}(90°)+E_3(p)\mathrm{S}(180°)+E_4(p)\mathrm{S}(270°) = E_2(p)-E_4(p) \quad (2)$$

$$\alpha(p) = \mathrm{At}(E_2(p) - E_4(p), E_1(p) - E_3(p)) \quad (3)$$

[1] The approach for computing the local edge orientation plays only a secondary role in the paper and therefore the proof of correctness is omitted here.

$$O_{rd}(p) = \begin{cases} \pi - (0.5 * \alpha(p)) & : \quad \alpha(p) > 0 \\ -(0.5 * \alpha(p)) & : \quad \alpha(p) \leq 0 \end{cases} \tag{4}$$

$$O_a(p) = \mathrm{round}((O_{rd}(p)/\pi) * 180) \tag{5}$$

The specific choice of filter orientations reveal considerable simplifications in (1) and (2). Figure 1 (middle) shows the edge orientation $O_a(p)$ for each point of the image I_g. The degrees are encoded as grey values reaching from black (0 degree) to white (180 degree).

Parallel to this procedure we compute the gradient magnitude of the image patch, detect edges by setting a threshold, and thus compute a binary image I_b (see Figure 1, right). Let $X_o \subset X$ be the subset of those coordinate tupels p where the gradient magnitude surpasses the threshold. Then we have the definiton $I_b(p) = 1$ for every $p \in X_o$ and $I_b(p) = 0$ elsewhere.

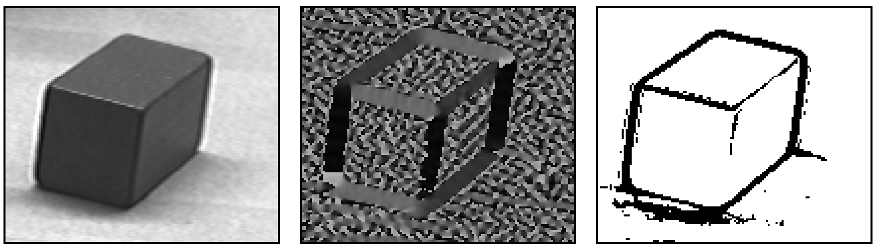

Fig. 1. Subimage with target object (left), orientation image for the edges (middle), binary image after setting a threshold to gradient magnitudes (right).

Second step: *Computing orientation-selective Hough transformation*
It is easily proven that edge points which lie along a line satisfy

$$L(p,q) := x1 * \mathrm{C}(\phi) + x2 * \mathrm{S}(\phi) - r, \qquad L(p,q) = 0 \;, \tag{6}$$

where $p = (x1, x2)$ and $q = (r, \phi)$. The two parameters r and ϕ of a two-dimensional real parameter space Q specify a line as follows: Parameter r is the distance from the image center to the line along a direction normal to line. The value reaches from $\frac{-id}{2}$ to $\frac{+id}{2}$, where id is the length of the image diagonal. Parameter ϕ is the angle of this normal direction to the $x1$ axis and reaches from 0 to 180 in angle degrees. We prefer this line representation because no singularities arise for vertical or horizontal lines.

The real parameter space Q is partitioned into a set $\overline{Q}$ of bins and accordingly each bin $\overline{q} \in \overline{Q}$ specifies a collection of lines with similar positions r and orientations ϕ. In our case we have a rectangular partition of bins with id vertical stripes and 180 horizontal stripes. The Hough transformation counts for each bin how many edges in the image lie along the lines specified by the bin.

Definition 1: The *Standard Hough Transformation* $SHT : \overline{Q} \rightarrow [0, \infty)$ of the binary image $I_b : X \rightarrow \{0,1\}$ relative to the line representation $L : X \times Q \rightarrow [0, \infty)$ is defined by

$$SHT(\overline{q}) = \#\{p \in X \mid I_b(p) = 1, \; and\ for\ some\ q \in \overline{q}\ holds : L(p,q) = 0\} \tag{7}$$

For reasons of discretization each bin is encoded by a single point in the parameter space. Therefore the bin array will be approximated by a matrix (the *Hough image*) of *id* columns and 180 rows and a point $q = (r, \phi)$ is defined by the index r of the column and the index ϕ of the row.

Figure 2 shows the Hough image resulting from the SHT of the binary image I_b in Figure 1 (right). We are interested in a Hough image with sharp peaks in order to easy locate the peaks and thus reliable estimate the line parameters. Unfortunately Figure 3 which depicts a selected patch of the Hough image in three dimensions (see rectangle in Figure 2) shows wide-spread maxima from which it is difficult to detect the relevant peaks. Obviously, all edges near or on a line cause the Hough transform to not only increase the content of the relevant bin but also many in its neighborhood.

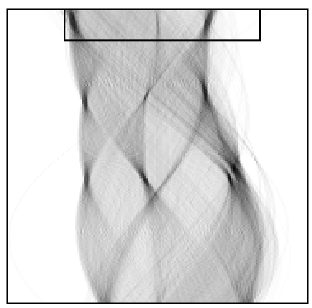

Fig. 2. Standard Hough Transform.

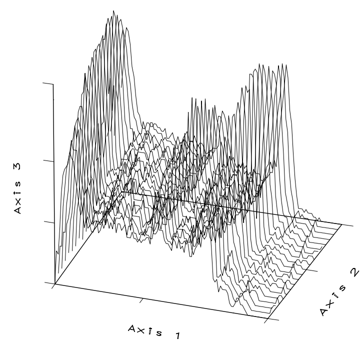

Fig. 3. SHT, top patch in 3D.

It's true making the partitioning of the parameter space more fine-grained would improve the accuracy, but in this case the performance time for computing the Hough transform increases. In order to avoid this trade-off we make use of the orientation image. The idea is to use the orientation $O_a(p)$ at an edge point and accumulate only the appropriate bin and furthermore those bins in a small neighborhood which are determined by a tolerance band.

Proposition 1: The orientation ϕ of a line is approximately identical to the orientation $O_a(p)$ of an edge point $p = (x1, x2)$ on the line, if the edge point arises from a 3D boundary line of the object:

$$x1 * \mathrm{C}(O_a(x1, x2)) + x2 * \mathrm{S}(O_a(x1, x2)) - r \approx 0 \tag{8}$$

Due to inaccuracies in the imaging process and the computation of edge orientation we include a *tolerance band* $\triangle a$ for O_a.

Definition 2: The *Orientation-selective Hough Transformation* $OHT : \overline{Q} \rightarrow [0, \infty)$ of the binary image $I_b : X \rightarrow \{0, 1\}$ relative to the line representation $L : X \times Q \rightarrow [0, \infty)$ is defined by

$$\begin{aligned} OHT(\overline{q}) = \#\{p \in X \mid I_b(p) = 1, and\ for\ some\ q = (r, \phi) \in \overline{q}\ holds: \\ L(p, q) = 0,\ and\ \phi - \frac{\triangle a}{2} \leq O_a(p) \leq \phi + \frac{\triangle a}{2}\}\ . \end{aligned} \tag{9}$$

Figures 4 and 5 show the effect of the OHT for the top patch in Figure 2. In order to realize the improvement with regard to sharpening the peaks the profiles must be compared to the SHT in Figure 3. Two examples of tolerance bands for the edge orientations has been used, $\triangle a = 9$, and $\triangle a = 3$ angle degrees. The result is that the smaller the tolerance band the sharper the peaks, and therefore the value of $\triangle a$ proves to be important.

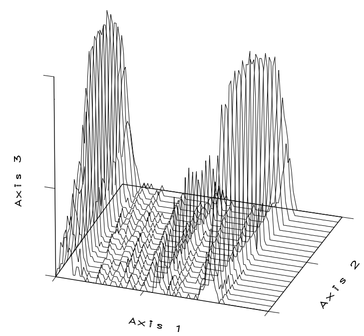

Fig. 4. Orientation-selective Hough Transform, large tolerance band for edge orientation.

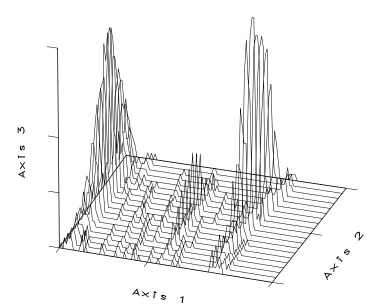

Fig. 5. Orientation-selective Hough Transform, small tolerance band for edge orientation.

Each peak in the Hough image of the OHT defines the parameters of a line of grey level edges in the image. This is the basis for building the boundary line configuration (BLC) of the target object. We are concerned with two problems. First, only those lines are of interest which arise from the surface of the target object. Unfortunately, there are many imponderables like shading, surface texture, partial occlusion, etc. causing spurious lines or deletion of essential lines from the object surface. Second, each peak in the Hough image represents an unbounded line. But we are interested in the bounded line segments which belong to the BLC of the target object. The first problem is handled in the next sections, for the second problem we refer to the work in [5].

3 Projective quasi-invariants of parallelepipeds

In this situation more cognitive information has to meet the low level process. Concretely, in order to extract the relevant peaks in the Hough image we make use of knowledge about the scene objects for which the BLC is requested. As opposed to using specific models for certain target objects we consider geometric invariants for general object shapes. We show exemplarily how to incorporate projective quasi-invariants of parallelepipeds for computing their BLCs. For purpose of demonstration the rectangular solid (shown in Figure 1, left) is used, which is a specific parallelepiped.

It is well known for the perspective projection of an ideal pinhole camera that an image point $(x1, x2)$ from a 3D scene point $(y1, y2, y3)$ is computed by

$$x1 = f\frac{y1}{y3}\ , x2 = f\frac{y2}{y3}\ , \tag{10}$$

where f is the distance between the center of perspectivity and the projection plane. Furthermore, it is known that parallel lines in 3D are no longer parallel after perspective projection into the image according to the equations of (10).[2] Fortunately, for certain imaging conditions the parallelism is *quasi-invariant* under perspective projection.

Definition 3: Let δ be the allowed deviation from parallelism in angle degrees.[3] Two lines with the values ϕ_1 and ϕ_2 of the angle parameter are *δ-approximate parallel*, iff

$$||\phi_1 - \phi_2|| \leq \delta \tag{11}$$

The parallelism is a *δ-quasi-invariant* iff parallel lines in 3D are δ-approximate parallel after perspective projection.

Notice that in the Hough image the bins of a row indicate the occurrence of parallel image lines of a certain orientation.

Proposition 2: For imaging conditions, in which the parallelism is a δ-quasi-invariant of the perspective projection, parallel lines in 3D occur as peaks in the Hough image being located within a horizontal stripe of height δ. The peaks in a stripe indicate δ-approximate parallel lines in the grey level image.

To get an impression for that, we describe the imaging condition for taking the picture in Figure 1, left). The distance between the camera and the target object was about $400mm$, and the lense of the objective was of $12mm$ focal length. Figure 6 shows nine peaks in the Hough image each one indicating a boundary image line of the target object. We realize three groups in horizontal stripes and each stripe contains three peaks. The vertical variation of the peaks in a horizontal stripe gives the deviation from parallelism. For the mentioned imaging condition it is at most 15 angle degrees.

In order to reliably extract the relevant peaks as depicted in Figure 6 we make use of the parallelism quasi-invariant. This leads to extracting the BLC of the rectangular solid.

4 Extracting the BLC of parallelepipeds

First step: *Emphasizing Hough peaks using the edge orientation histogram*
In section 2 the orientation-selective Hough transformation has been introduced and the important role of the tolerance band for edge orientations was demonstrated. In order to reduce the role of this tolerance band and/or to sharpen the peaks in the Hough image nevertheless we apply a further transformation.

According to proposition 1 there is a consensus between the orientation of a grey level edge and the orientation of the line through the edge point. Furthermore it was stated in proposition 2 that parallel object lines are quasi-parallel

[2] Except for lines parallel to the projection plane.

[3] The user must set this parameter depending on the shape of the target objects.

in the image. Therefore, if we count the grey level edges of a certain orientation we get the number of points lying on parallel image lines of this orientation. On the other hand in the Hough image a certain row indicates a certain orientation of parallel image lines. Therefore, by multiplying a Hough row with the number of grey level edges of the orientation specified by the index of the row, we can more or less emphasize the bins of this row.

According to that we compute a global histogram of edge orientations from all points in the binary image (in Figure 1, right). The histogram counts the number of edges respective for the orientations and is defined in the interval D between 0 and 180 angle degrees. Figure 7 shows the Edge Orientation Histogram for the image in Figure 1 (middle).

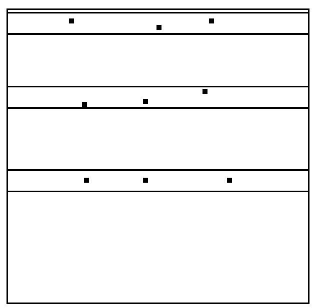

Fig. 6. Three groups of peaks in horizontal stripes.

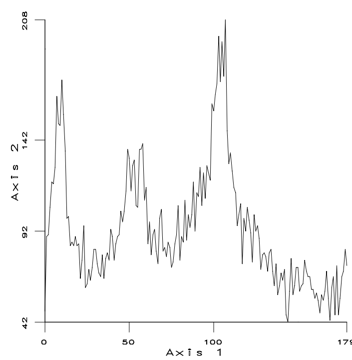

Fig. 7. Edge Orientation Histogram.

Definition 4: Let $\overline{D}$ be the partition of D into 180 unit intervals and $\overline{d} \in \overline{D}$. The *Edge Orientation Histogram* $EOH : \overline{D} \rightarrow [0, \infty)$ of the orientation image $O_a : X \rightarrow [0, 180]$ is defined by

$$EOH(\overline{d}) = \#\{p \in X \mid O_a(p) \in \overline{d}\} \tag{12}$$

For combining the OHT with the EOH we realize that the Hough image consists of 180 rows and the histogram is a vector of 180 components. A row of index ϕ in the Hough image contains hints for the occurrence of image lines of orientation ϕ and the component of index ϕ of the histogram vector contains the number of edges having orientation ϕ. In order to weight the row values of a certain orientation by the number of edges of the same orientation we transform the histogram vector into a diagonal matrix and multiply it with OHT.

Definition 5: The Weighted Orientation-selective Hough Transformation (WOHT) is defined by

$$WOHT = Diag(EOH) * OHT, \tag{13}$$

where $Diag(EOH)$ is a diagonal (180×180)-matrix whose diagonal elements are the vector components of the Edge Orientation Histogram (EOH), and OHT is the $(180 \times id)$-matrix of the Hough image.

Figures 8 and 9 show the effect of the WOHT by taking as tolerance band for the edge orientations $\triangle a = 9$ and $\triangle a = 3$ angle degrees. We realize that the role of this parameter has been reduced significantly because both results look similar. A result is reached which is similar (or even better) to applying OHT with $\triangle a = 3$ angle degrees (compare Figures 8 and 9 with Figure 5).

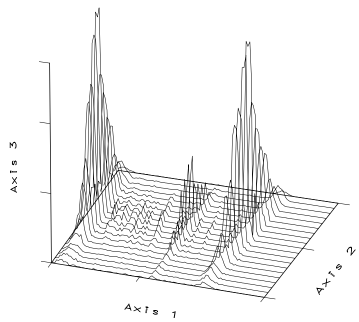

Fig. 8. WOHT, large tolerance band.

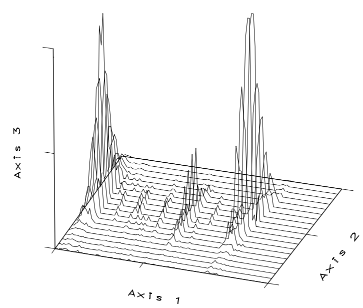

Fig. 9. WOHT, small tolerance band.

Second step: *Peak detection by using the parallelism quasi-invariant*
In the case of a parallelepiped one typically looks at three plane faces, and these are represented by three classes of line segments, where each class consists of three δ-approximate parallel lines. Based on this insight the search of the relevant peaks is easy. We are searching for the global maximum peak and thus determine the first relevant horizontal stripe. Within the stripe two other maximum peaks must be detected. Then the stripe is erased completely and in this modified Hough image we look for the next global maximum. This defines the second relevant stripe in which we once again detect two further maximum peaks. Repeating the procedure for the third stripe we finally have nine relevant maxima as shown in Figure 6. The lines belonging to the maxima are depicted in Figure 10 on top of the greylevel image. We realize for each line the touch with a part of the object boundary.

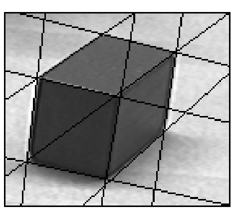

Fig. 10. Lines for rectangular solid.

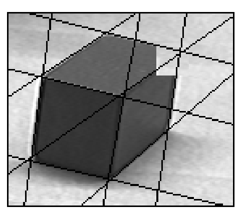

Fig. 11. Lines under solid occlusion.

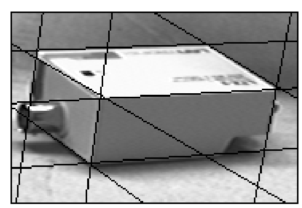

Fig. 12. Lines for a complicated object.

5 Summary and Discussion

We presented an approach for extracting line configurations from an image to describe the boundary of objects. The standard Hough transform for estimating line parameters has been modified with the purpose of sharpening the peaks. For this the consensus principle of edge-line-orientation was taken into account. The perspective quasi-invariant of parallelism is used for locating the appropriate set of Hough peaks which describes the boundary lines of an individual object.

Using these fundamental principles we succeeded in extracting relevant image structures even in scenes for which specific object models are not available. The approach also works if the rectangular solid is under partial occlusion (simulated in Figure 11). More complicated objects can be handled as well, e.g. objects whose shape partially differ from a parallelepiped and additionally may have a non-homogeneous greylevel structure on the surface (see Figure 12).

The work should be continued by making use of other geometric invariants which are valid for more complicated objects. An example is the incidence invariance of perspective projection, e.g. the intersection points of 3D object lines must transform in crossing points of image lines. In [5] it is shown how to use this principle to fine-tune the line parameters. Further photometry/geometry consensus principles must be realized, for example to determine especially those segments of the lines which build up the boundary line configuration of the object (see again [5]). Finally, criteria for testing the validity of invariances and consensus principles are needed, e.g. more detailed plausibility criteria for the set of located peaks in the Hough image. This is useful if we have a degenerate view of an object or interior boundary lines are not available due to missing greylevel contrast between the object faces.

Acknowledgement: The discussions with Gerald Sommer are appreciated.

References

1. Jähne, B.: Digitale Bildverarbeitung. Springer Verlag, Berlin (1993) 128-138
2. Jones, J., Lozano-Perez, T.: Planning two-fingered grasps for pick-and-place operations on polyhedra. IEEE Int. Conf. on Robotics and Automation (1990) 683–688
3. Kollnig, H., Nagel, H.-H.: 3D pose estimation by fitting image gradients directly to polyhedral models. Int. Conf. on Computer Vision (1995) 569–574
4. Leavers, V.: Survey - Which Hough transform ? Computer Vision and Image Understanding **58** (1993) 250–264
5. Pauli, J.: Geometry/Photometry consensus for object boundary extraction. submitted to ICCV (1997)
6. Princen, J., Illingworth, J., Kittler, J.: A hierarchical approach to line extraction based on the Hough transform. Computer Vision, Graphics, and Image Processing **52** (1990) 57–77
7. Zisserman, A., Mundy, J., Forsyth, D., Liu, J.: Class-based grouping in perspective images. Int. Conf. on Computer Vision (1995) 183–188