Probabilistic Decisions in Production Nets: An Example from Vehicle Recognition

E. Michaelsen and U. Stilla

FGAN-FOM Research Institute for Optronics and Pattern Recognition
Gutleuthausstr. 1, 76275 Ettlingen, Germany
mich@fom.fgan.de, stilla@fom.fgan.de
http://www.fom.fgan.de

Abstract. A structural knowledge-based vehicle recognition method is modified yielding a new probabilistic foundation for the decisions. The method uses a pre-calculated set of hidden line projected views of articulated polyhedral models of the vehicles. Model view structures are set into correspondence with structures composed from edge lines in the image. The correspondence space is searched utilizing a 4D Hough-type accumulator. Probabilistic models of the background and the error in the measurements of the image structures lead to likelihood estimations that are used for the decision. The likelihood is propagated along the structure of the articulated model. The system is tested on a cluttered outdoor scene. To ensure any-time performance the recognition process is implemented in a data-driven production system.

1 Introduction

Vehicle recognition from oblique high resolution views has been addressed by several authors [2][7][6]. Hoogs and Mundy [7] propose to use region and contour segmentation techniques and rely on dark regions of certain size and form, that may be a vehicle shadow, and on simple features like parallel contours, that some vehicles display in a variety of perspectives. Shadows can be exploited, if the pictures are taken in bright sunlight of known direction. Omni-directional ambient lighting causes a shadowed region directly underneath the vehicle. This is visible in oblique views of vehicles but may be occluded, e.g. by low vegetation. Parallel contours are a cue to vehicles, but they are present in many environments around vehicles, too (e.g. in roads, buildings, ploughed fields).

A possibility to avoid this difficulties is to use the geometrical shape of the vehicles themselves. Viola and Wells [12] render object models and compare characteristic properties of the gray value function of the rendered graphic and the image using mutual information. Hermitson et al. [6] utilize this approach to oblique vehicle recognition. Rendering requires assumptions about the lighting and surface properties of the model. If this is not available one has to work with contours on the more abstract geometric level. Dickinson et al. [3] proposed generalized cylinder models with part-of hierarchies for contour based object recognition. Binfort and

Levitt [2] applied this to vehicle recognition tasks. Generalized cylinder models capture the coarse structure of a vehicle. For details of vehicles such models are not appropriate.

Grimson [5] proposed polyhedron models and straight line segments. This has a high potential discriminative power, because many geometric properties and constraints of the targets are exploited. For the reduction of the computational effort indexing methods like generalized Hough transform as well as restricting the vehicles in position and rotation to the ground plane are proposed [10]. Some vehicles can not be covered by one rigid polyhedron alone, because they are composed of parts, that are connected by pivots or bearings (e.g. truck and trailer systems or tanks). Such objects can be captured by articulated models [11]. The appearance of polyhedrons is affected by self occlusion. This may be treated by aspect graphs [4], or by linear combination of characteristic views [11]. We use an equidistantly sampled set of views for each model [8]. In this contribution we incorporate probabilistic calculations into a structural approach.

Sect. 2 presents the accumulator method to solve the problem of vehicle recognition from single oblique views. The probabilistic model is described in Sect. 3. A result of an experiment on a difficult scene is given in Sect. 4. In Sect. 5 a discussion of pros and cons of the approach and an outlook on future work are given.

2 View-Based Recognition of Vehicles

View-based object recognition matches the model to the data in the 2D image space. For this purpose 2D views of the 3D model parts are constructed. It is possible to use structured models with part-of hierarchies. Then the consistency check for correct mutual positioning requires back projection.

A set of 2D lines constructed by perspective hidden-line projection from a polyhedron is called a view. In contrast to this an aspect is a line graph. Changes in the view that don't change the topology provide the same aspect [11].

2.1 The Space of views

The space of views is originally continuous and has dimension six (three rotations and three translations). Vehicle recognition from oblique imagery constraints the distance to an interval and the spatial rotation to one off-image plane rotation (the azimuth). Depending on the focal length translations of the model may lead to geometric distortions at the margins of the image. Due to the long focal lengths used here this effect can be neglected and the same view model can be used all over the image. The model is positioned such that it appears centered in the principal point and the azimuth and distance are varied stepwise in an appropriate step width yielding finite 2D view space containing some hundred views per model. Fig. 1 shows some example views.

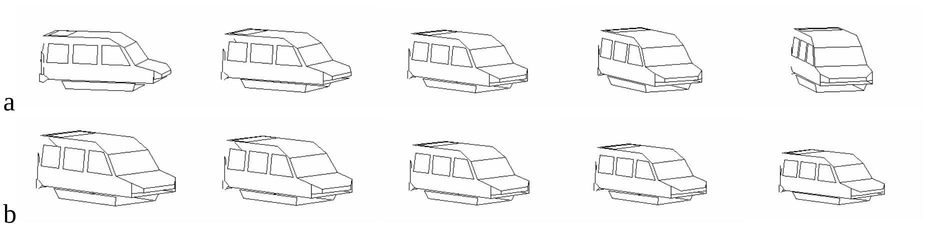

Fig. 1. Selected set of 2D models projected from a 3D polyhedron model: a) varying azimuth with $\Delta\alpha=15^o$; b) varying distance with $\Delta dis=8m$

2.2 Matching an Image to the Views

Object contours in the image are extracted using a gradient operator and morphologic thinning. The contours are approximated by short line segments. A line prolongation process improves the orientation estimation of the line objects. The set of such line objects can be matched with the lines in the views. For this task we use a generalized Hough transformation [1].

To decrease the computational complexity of the correspondence search we use L-shaped objects constructed from the lines. The L-shaped objects in all the model views are constructed off-line. As key to establish the correspondence between image and model structures the two orientations of the sides of the L-shaped objects are utilized. A structure in the image supports a part of a view if both orientations are sufficiently similar. The position of the reference point of the view is obtained subtracting the position of the part in the model view from the position in the image.

2.3 Robustness through Accumulation

Often not all modeled structures are present in images of outdoor scenes. Therefore, as much evidence as possible has to be merged from consistent cues to one specific pose. While a single cue may result from background or clutter multiple consistent cues from different structures of a specific view probably result from the presence of the modeled object in the corresponding pose. Therefore all cues are inserted in a 4D accumulator at their image position, azimuth, and distance. Resulting from different errors (modeling, imaging, feature extraction) consistent cues form a fuzzy cluster in the accumulator. For the detection of vehicles we search for dominant clusters of cues in the accumulator.

Each cue locates a 4D search area. The size of this area results from the maximal expected errors. Cues within a search area are a candidate subset for a cluster object and are used to estimate the center of mass. The center of mass locates a new search area and a new subset. Such

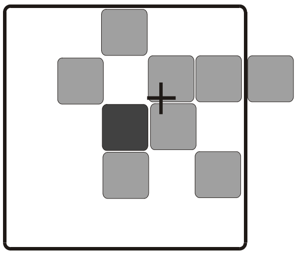

Fig.2. Searching for a proper subset in the accumulator

calculations are performed until convergence occurs. Fig. 2 exemplifies such procedure in 2D where the dark square indicates the position of a cue and the black square shows the corresponding search area. While the leftmost cue is missed in the first attempt it will be included in a later step, because the position of the new search area is determined by the center of mass indicated by the cross.

2.4 Part-of Hierarchies and Articulated Models

Not all vehicles are adequately described by a single shape fixed polyhedron model. Parts of a vehicle may be mutually connected and constrained by hinges or pivots (truck-trailer systems, tanks). Therefore we consider 3D models of vehicles that have a part-of hierarchy. Such a model is described by a directed graph where each basic part is a polyhedron. If the parts have mutual degrees of freedom in rotation such a model is called articulated model [11]. The resulting constraints are used by recognition process. For the consistency test the parts are projected back to the 3D scene. If a pivot or hinge is not located at the reference position of a model part, then auxiliary position attributes are used to define the search areas for partner clusters. E. g., the 2D position of the trailer hitch of a vehicle view depends on its pose. These auxiliary position attributes locate the search area for possible partners.

The information on which auxiliary attribute of which part of the model connects to which attribute of which other part, and which azimuth angle differences are permitted at this connection is given by the user in a standardized format in addition to the polyhedron models.

2.5 Production Nets and Implementation

We describe structural relations of the object models by productions. A production defines how a given configuration of objects is transformed into a single more complex object (or a configuration of more complex objects). In the condition part of a production geometrical, topological, and other relation or attributes of objects are examined. If the condition part of a production holds, an object specific generation function is executed to generate a new object. Such productions operate on sets of objects instead of graphs, strings etc. The organization of object concepts and productions can be depicted by a production net [9] which displays the part-of hierarchies of object concepts.

Our production nets are implemented in a blackboard architecture. Blackboard-systems consists of a global data base (blackboard), a set of processing modules (knowledge sources), and a control unit (selection module). The productions are implemented in the processing modules, which test the relations between objects and generate new objects. Starting with primitive objects the searched target objects are composed step by step by applying the productions. The system works in an accumulating way, this means a replaced initial configuration will not be deleted in the database. Thus all generated partial results remain available during the analysis to pursue different hypotheses. The classical backtracking in search-trees is not necessary.

3. Probabilistic Error Models

A critical issue is the choice of the optimal size of the search areas in the accumulator. With rising distance of a cue from the center of a cluster the likelihood for its membership decreases. A cue with a large distance from the cluster is probably due to background or clutter. Wells [14] used Gaussian distributions for the error of features that are in correct correspondence to the model and equal densities for background and clutter features. While he uses contour primitives attributed by their location, orientation and curvature we operate in the 4D accumulator.

3.1 Probabilistic Calculations in the Cluster Formation

Applying Wells theory we first have to estimate a reward term λ as contribution of each single cue which replaces the entry into the accumulator. From a representative training-set where the features are labeled either as correctly matched or as background or prior information λ is set to

$$\lambda = \ln\left(\frac{1}{(2\pi)^2 m} \cdot \frac{(1-B)}{B} \cdot \frac{W_1 \cdots W_4}{\sqrt{|\psi|}} \right). \tag{1}$$

The middle factor in this product is calculated from the ratio between the probability B that a feature is due to the background, and the probability $(1-B)/m$ that it corresponds to a certain model feature, where m is the number of features in the model. The rightmost factor in the product is given by the ratio between the volume of the whole feature domain W_1 ... W_4 and the volume of a standard deviation ellipsoid of the covariance matrix ψ for the correctly matched features. As feature domain we set $\beta^T=(x,y,\alpha,dis)$. Locally our accumulator domain may be treated as linear, justifying the application of this theory and its error models. The objective function L is calculated for each cluster of cues:

$$L = \sum_j \left[\lambda - \operatorname*{Min}_{\Gamma_i = j} \left[\frac{1}{2} (Y_i - \hat{\beta})^T \psi^{-1} (Y_i - \hat{\beta}) \right] \right]. \tag{2}$$

Y_i is the position of the i-th cue in the accumulator domain. The pose β is estimated as mean $\hat{\beta}^T = (\hat{x}, \hat{y}, \hat{\alpha}, \widehat{dis})$ of the poses of the member cues of the cluster. The correspondence Γ is coded as an attribute of the cues. For each model feature j put into correspondence in the cluster the closest cue i to the mean is taken as representative of the set of all cues i corresponding to j. This is done, because we regard multiple cues to the same model feature as not being mutual independent.

Recall that the maximization must not take those Γ into account, that include negative terms into the sum. Fig. 4 displays the 1D case: Full reward λ is only given for a precise match. With rising error the reward is diminished by a negative parabola. Finally it reaches zero level. At this point Γ is changed, setting the feature in correspondence to the background. This condition gives a new way to infer the threshold parameters for the search region in the cluster process. In 1D the covariance matrix reduces to a single variance σ and the single threshold parameter *d* is given by the root of λ/σ. For higher dimensional cases (e.g. 4D) the bounding box of the ellipsoid is used, that is determined by the covariance Σ and λ.

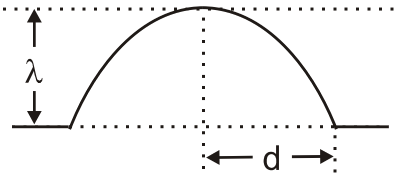

Fig. 3. Reward function after Wells [11]

Wells rejects scenes as non recognizable, if λ turns out to be negative according to Eq. 1. This gives a profound criterion for the applicability of the approach to a task for which a test data set is provided.

3.2 Propagation of likelihood along the part-of structure

The cues have an auxiliary attribute Y_1 for the position, where the partner cue should connect, e.g. the trailer hitch. This attribute is calculated by inverse projection into the scene, proper rotation of the 3D model, and projection into the image again. A search area is constructed around Y_1. For each partner cue with position Y_2 in this area the aggregation is regarded as valid and the two accumulator values are summed up yielding the accumulator value for the new aggregate object. Its position Y_n is calculated as weighted mean. This neglects the quality of the fit.

For the probabilistic setting the likelihood L is propagated along the links of the part of hierarchy. If the position Y_2 of the partner cue exactly matches the auxiliary position Y_1, we infer that there is independent evidence for the aggregate from both parts. This justifies multiplication of probabilities or adding the likelihood values. Otherwise some of the predecessors of the cue clusters may be contradicting. Lacking the precise knowledge of the distribution, the evidence for each part is assumed to be equally distributed over its search volume. Fig. 4 shows the 1D case.

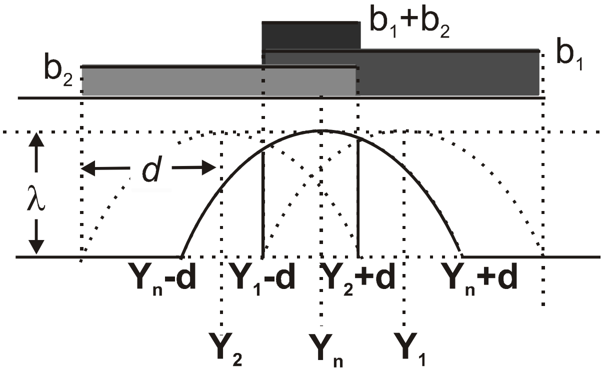

Fig 4: combining evidence from two different parts of a model into evidence of an aggregate: Below reward function; above density estimations.

The two cue clusters with centers Y_1 and Y_2 and their error parabolas displayed in dashed lines include mutually affirming evidence, if their distance is smaller than *3d.*

We indicate their evidence densities b_1 and b_2 by the differently shaded piecewise constant functions. In the overlapping region evidences b_1, b_2 are added. The reward function integrates the piecewise densities using the error parabola of the new position estimate Y_n yielding a sum of three integrals:

$$L_n = \int_{Y_n-d}^{Y_1-d} b_2 (x - Y_n)^2 \sigma \, dx + \int_{Y_1-d}^{Y_2+d} (b_1 + b_2)(x - Y_n)^2 \sigma \, dx + \int_{Y_2+d}^{Y_n+d} b_1 (x - Y_n)^2 \sigma \, dx \quad (3)$$

In the 2D case the upper and lower border of the integral in the middle are replaced by circular sections in the attribute domain and the parabola is replaced by a paraboloid. In the case of rigid connections the hyper ellipsoid on which the reward paraboloid is constructed is 4D, namely *(x,y,*α*,dis)*. In the case of an articulated connection the azimuth α is free contributing no error. Therefore the domain is 3D containing only *(x,y,dis)*.

4 An Experiment

Fig. 5a shows a section of a gray level image containing a vehicle with a small trailer. From this lines are extracted (Fig. 5b). Experiments very carried with this data. Both decision criteria, the maximum accumulator value as well as the maximum likelihood (ML) work. In the cluttered image region on the left (branches of a tree) and in the fairly homogenous region in the center accumulator and likelihood field are empty.

The pose estimation of the maximal elements is roughly correct. Fig. 5c shows the ML result. The interesting section of the likelihood field is enlarged in Fig 5d. The white blobs on the left correspond to correct localization. Some less significant false evidence is found on the right. Both the discrimination and the pose estimation are slightly better for the likelihood criterion.

5 Discussion

In this contribution we demonstrated the inclusion of probabilistic calculations into a structural method. Compared to previous experiments [8] the discriminative power of the accumulator on cluttered regions, e.g. in the left part of the image, has much improved due to a better parameter setting. The new settings were obtained from the probabilistic considerations. We occasionally experienced better performance of the accumulator compared to the likelihood. This occurred when the model did not fit exactly to the vehicle. The likelihood approach is more sensitive to errors in the model. Fig. 5c shows that the pose is not optimal (see nose of the vehicle). EM type optimizations including a top down search in the correspondence space can improve the result [13].

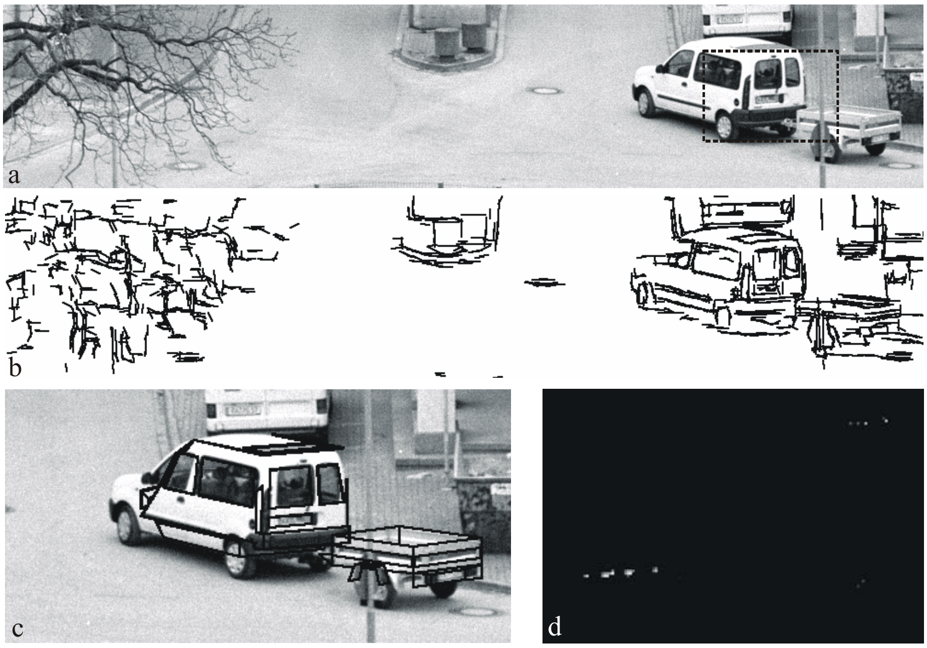

Fig. 5. Localization of an aggregate consisting of a vehicle and a small trailer. a) Image section (1000x200 pixel), b) extracted line objects; c) overlaid articulated model of ML-result, d) section of the likelihood field corresponding to the dashed box in a).

The probabilistic calculations of Wells rest on certain assumptions on the distribution of the data. Background features are assumed to be equally distributed all over the picture. Such assumption is valid only in special situations or if nothing else is known about the background [9]. If additional information is given, e.g. on certain preferences on the orientations of the lines (e.g. vertical or horizontal), this can be included in the probabilistic model. The features in correspondence to the target are modeled with a Gaussian distributed additive error. If knowledge about the error sources is available, other error models may be considered.

As shown in Fig 5 the evidence for the two partners of an aggregate is estimated as being equally distributed over the search volume. The evidence for the new aggregate has a stepwise constant density (lower, high and then lower again). If we include such an aggregate as a part of a higher aggregate using the same calculations, we permit a systematic estimation error. For shallow hierarchies like the one presented here this is not important. For deep hierarchies such effect has to be estimated.

In our approach all possible model views are approximated by views valid for the principal point only. This is justified for long focal lengths but will pose severe problems for views near the image margin of wide angle pictures. These are distorted by systematic errors.

Still the preliminary experiments presented in chapter 4 yielded promising results, so that we are confident in combining statistical and structural methods.

References

1. Ballard D. H., Brown C. M.: Computer Vision. Prentice Hall, Englewood Cliffs, New Jersey, (1982).
2. Binfort T. O., Levitt T. S.: Model-based Recognition of Objects in Complex Scenes. In: ARPA (ed.). Image Understanding Workshop 1994. Morgan Kaufman, San Francisco (1994), 149-155.
3. Dickinson S. J., Pentland A. P., Rosenfeld A.: From Volumes to Views: An Approach to 3-D Object Recognition. CVGIP:IU, Vol. 55, No. 2 (1992), 130-154.
4. Eggert D. W., Bowyer K. W., Dyer C.R.: Aspect Graphs: State-of-the-Art and Applications in Digital Photogrammetry. ISPRS-XXiX, Vol. 5, Com V, (1992) 633-645.
5. Grimson W. E. L.: Object Recognition by Computer: The Role of Geometric Constraints. MIT Press, Cambridge, Mass., (1990).
6. Hermitson K. J., Booth D. M., Foulkes S. B., Reno A. L.: Pose Estimation and Recognition of Ground Vehicles in Aerial Reconnaissance Imagery. ICPR 1998, Vol. 1, IEEE, Los Alamitos, California, (1998), 578-582.
7. Hoogs A., Mundy J.:An Integrated Boundary and Region Approach to Perceptual Grouping. ICPR 2000, Vol. 1, IEEE, Los Alamitos, California, (2000), 284-290.
8. Michaelsen E., Stilla U.: Ansichtenbasierte Erkennung von Fahrzeugen. In: Sommer G., Krüger N., Perwas C. (eds.): Mustererkennung 2000, Springer, Berlin, (2000) 245-252.
9. Michaelsen E., Stilla U.: Assessing the Computational Effort for Structural 3D Vehicle Recognition. In: Ferri F.J., Inesta J.M., Amin A., Pudil P. (eds): Advances in Pattern Recognition (SSPR-SPR 2000), Springer, Berlin, (2000) 357-366.
10. Tan T., Sullivan G., Baker K.: Model-Based Localisation and Recognition of Road Vehicles. . Int. Journ. of Comp. Vision, Vol. 27, (1998) 5-25.
11. Wang P. S. P.: Parallel Matching of 3D Articulated Object Reccognition. Int. Journ. of Pattern Recognition and Artificial Intelligence, Vol. 13, (1999) 431-444.
12. Viola P., Wells W. M. III: Alignment by Maximalization of Mutual Information. Int. Journ. of Comp. Vision, Vol. 24, (1997) 137-154.
13. Wells W. M. III: Statistical Approaches to Feature-Based Object Recognition. IJCV, 21, (1997) 63-98.