How Context And Ordering Constraints Can Improve 3D Object Recognition

Mike Födisch and Thorsten Hermes

TZI-Report No. 25, 2002

How Context And Ordering Constraints Can Improve 3D Object Recognition

Mike Födisch and Thorsten Hermes

Center for Computing Technologies, Image Processing Department,
P.O.B. 33 04 40, 28334 Bremen, Germany
{fmike, hermes}@tzi.org

Zusammenfassung We show how context and ordering information can improve the efficiency and accuracy of 3D object recognition. First we describe a way how to obtain contextual information in terms of features enriched by attributes. On this basis we introduce new constraints for the application of interpretation tree search. Finally, we present exemplary recognition results which show the impact of the new constraints.

1 Introduction

Recognition of three-dimensional (3D) objects is still an ongoing topic of interest in computer vision. Different approaches were developed to recognize objects on 3D image data (e.g. [FJ91]) or 2D image data (e.g. [BJ01]). Commonly, the first step of a recognition process consists of deriving significant features from image data. Subsequently, these features are matched to model features. Other approaches are based on learning the appearance of objects, e.g. probabilistic approaches [HN00] or neural networks [SD99]. In this work we focus on properties of features and the functional role of these properties for 3D object recognition. Therefore, we distinguish between — what we call — *pure* features and features *enriched* by attributes. Assuming that the attributes enhance the significance of the features, we postulate that the use of enriched features improves the efficiency of object recognition.

Our application domain uses the recognition of 3D polyhedral objects via interpretation tree search [Gri90]. This approach served as a basis for several systems which adjusted some aspects, e.g. ordering of features [AA93] or creating derived features [FJ91], [SM92]. We show how to obtain attributed features with context and ordering information (section 2). Furthermore, we use this information to formulate new constraints as needed for the interpretation tree approach to recognize objects (section 3). In section 4 we present exemplary recognition results which we gathered from testing the system against real data.

2 Obtaining contextual information

Contextual information is strongly bound to the feature where it belongs to. So, first there is a need for extracting features from input data. To extract edge-based features so-called gradient operators (see [JHG99]) are usually used. These features are then needed for further processing like object recognition.

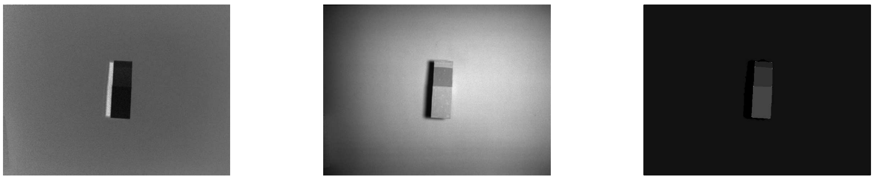

Abbildung 1. Range map, intensity image and segmentation results

We suggest another way to extract edge-like features. Our basic material consists of a range image from a 3D laser sensor and additionally a grey value image of a CCD camera. Both images are correlated to each other. After applying image enhancement methods (e.g. noise reduction) we use a sophisticated segmentation method to extract segments. The segmentation method combines edge- and region-based approaches [PL90]. Concerning the two input images in figure 1(left and middle) the segmentation method extracts the segments shown in figure 1(right). The segments are represented by their contour and a function which approximates the bordered area. The contour consists of reduced contour points which represent mostly vertices of polyhedral objects. For the extraction of edge-like features we project lines between adjacent contour points with respect to clockwise ordering. These lines represent features which belong to a context (segment) and possess an inherent ordering.

3 Object Recognition

Our object recognition method is based on the hypothesize-and-test approach [Gri90]. First hypotheses are generated from some evidence and verified afterwards. The verification consists of global consistency checking and selecting the best hypothesis. Our concrete implementation represents a variation of the *interpretation tree search approach* introduced by Grimson [Gri90]. The interpretation tree is used to span the correspondence space which means that all possible combinations of associations between scene features and model features can be generated. At each level of the interpretation tree one scene feature is tested in order to associate it to all model features. If such an association is possible w.r.t to constraints, the next level of the tree is spanned and therefore the next scene feature is considered. If such an association is not allowed due to *constraints* (e.g. length check) the tree is pruned at that node and further associations starting from here are no longer considered. If all scene features are associated to model features, all paths starting from the root of the tree to a leaf represent interpretations of the scene. The interpretations vary regarding the quality of associated scene features (e.g. length of edge-like features). Only those interpretations which show the best quality are considered as *hypotheses*. The final

verification of the remaining hypotheses separates good and wrong hypotheses via computing the best transformation of model data into the scene pose. The one and only hypothesis which shows the lowest transformation error represents the ultimate interpretation of what is seen in the scene.
Our implementation differs from Grimson most notably in the type of features used and new introduced constraints. As described in section 2 our features consist of linear edge fragments with additional context information. On the basis of those features we introduce a new constraint in the following.

3.1 Context-Consistency Constraint

We introduced in [FWH01] the *Correspondence-Consistency Constraint* which claims the membership of model features to the same model surface in certain circumstances. We enhance this constraint to the *Context-Consistency Constraint* ($\mathcal{C}^3$). Besides the membership of model edges to model surfaces (context) this new constraint also claims the compliance of *ordering* of model edges. In this section we describe both aspects of $\mathcal{C}^3$ and finish with a complete formalism of the constraint.

3.1.1 Membership to context

The $\mathcal{C}^3$ pretends that scene features of the same segment have to refer to model features of the same model surface. The example in figure 2 illustrates this requirement. Given, that scene feature s_{13} was already associated to model feature m_3 and then an association for scene feature s_{12} should be found. There are two model surfaces F_1, F_2 adjacent to model feature m_3. They represent the context for further associations. $\mathcal{C}^3$ only allows associations of s_{12} to one of the model edges belonging to the context (thus $m_1, \ldots, m_7$). An association to other model edges would be not allowed because they do not belong to either model surface F_1 or model surface F_2. After associating the second scene feature s_{12} (e.g. to model feature m_2) the context consists of exactly one remaining model surface (F_1).

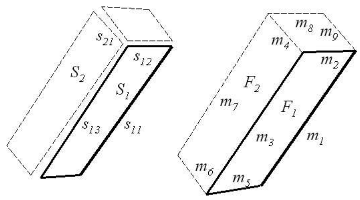

Abbildung 2. Membership aspect of the $\mathcal{C}^3$

3.1.2 Ordering in context

A second important aspect of $\mathcal{C}^3$ deals with the ordering of model features in associations.

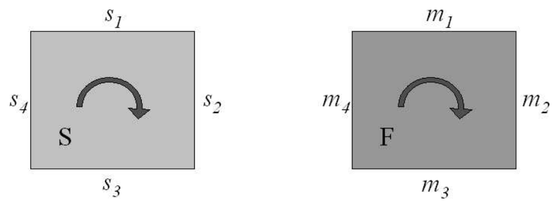

Abbildung 3. Ordering aspect of $\mathcal{C}^3$

Figure 3 shows a segment S, a model surface F and appendant scene features s_i and model features m_j. The scene and model features are ordered clockwise within the context they belong to (segment and model surface). Beginning at this starting point, the effect of $\mathcal{C}^3$ is described on the basis of the following examples:

- The new constraint allows the trivial *synchronized* association of scene features to model features. Therefore both of the following interpretations are allowed (see section 3.2 for notation):
  1. $I_4^a = \{(s_1 : m_1), (s_2 : m_2), (s_3 : m_3), (s_4 : m_4)\}$
  2. $I_4^b = \{(s_1 : m_3), (s_2 : m_4), (s_3 : m_1), (s_4 : m_2)\}$
- Starting from the interpretation $I_2^c = \{(s_1 : m_1), (s_2 : m_3)\}$ a further association $(s_3 : m_2)$ would be not allowed because m_2 is situated *before* the last associated model feature m_3 w.r.t. clockwise ordering. Nevertheless an association $(s_3 : m_4)$ would be allowed because m_4 is situated *behind* m_3 w.r.t. the ordering.
- Starting from the interpretation $I_2^d = \{(s_1 : m_1), (s_2 : m_4)\}$ no further associations are allowed because there are no more model features *behind* m_4 and *before* m_1 w.r.t. clockwise ordering. This example shows how the $\mathcal{C}^3$ can exclude mirror-inverted associations. Because of the local effect of constraints mirror-inverted and therefore wrong associations often occur using interpretation trees. The usage of contextual information in our approach prevents such hypotheses.

3.2 Formalism

In this section we describe the $\mathcal{C}^3$ in a more formal way. First, general notations are presented followed by formal notation of the two aspects of $\mathcal{C}^3$. Finally both parts will be combined for the complete notation.

A segment S consists of n linear edge-like features s_i:

$$S = \{s_1, s_2, \dots, s_n\} \quad \text{ordered set,} \quad \text{with} \tag{1}$$
$$s_i = (P_b, P_e), \tag{2}$$

and $P_b, P_e \in \mathbb{R}^3$ are three-dimensional points. A model surface F consists of f linear object edges m_j:

$$F = \{m_1, m_2, \dots, m_f\} \quad \text{ordered set,} \quad \text{with} \tag{3}$$
$$m_j = (P_b, P_e), \tag{4}$$

and $P_b, P_e \in \mathbb{R}^3$ are three-dimensional points. Additionally, we use a concept for the representation of an object. An object $\mathcal{O}$ consists of o model surfaces F_k:

$$\mathcal{O} = \{F_1, F_2, \dots, F_o\}. \tag{5}$$

An interpretation I consists of a set of associations $(s_i : m_j)$ between scene features s_i and model features m_j. Interpretations are generated as follows:

$$I_{n+1} = I_n \cup \{(s_{i_{n+1}} : m_{j_{n+1}})\}, \tag{6}$$

with $I_0 = \{\} = \emptyset$ for the interpretation without any associations. Furthermore, we introduce the following notation: $I_i.s$ determines the scene feature and $I_i.m$ determines the model feature of the i-th association in interpretation I.

The association of scene feature $s \in S$ to a model feature m of an object $\mathcal{O}$ leads to the following context C which holds for further associations:

$$C = \{F_l | m \in F_l\}, \quad \text{with } F_l \in \mathcal{O}, l \in \mathbb{N}^+ \tag{7}$$

with $\mathbb{N}^+$ set of the natural numbers *without* null.

For some object $\mathcal{O}$ the model context C_{I_n} of an interpretation I_n with $n > 0$ can generally be formulated as follows:

$$C_{I_n} = \bigcap_{i=0}^{n} \{F_l \in \mathcal{O} | I_i.m \in F_l\}, \quad \text{with } l \in \mathbb{N}^+ \tag{8}$$

We want to remark that the context C_{I_n} for interpretations I_n with $n > 1$ consists of exactly one remaining model surface F_C.

The claim for context consistency of $\mathcal{C}^3$ can thus be formulated as follows:

$$\mathcal{C}^{3a}_{I_n}((s : m)) \Leftrightarrow (n = 0) \vee (\exists F_l \in C_{I_n} : m \in F_l). \tag{9}$$

For each new association $(s : m)$ added to a given interpretation I_n with $n > 0$ has to exist a model surface F_l in context C_{I_n} which contains $m \in F_l$ as a model edge. If there was not yet an association in interpretation I_n with $n = 0$ the claim for context consistency is fulfilled in a trivial way.

The clockwise ordering is mapped onto the sequence of model edges m_j in a model surface F. As for the example in figure 3 the elements of F hold the following sequence $F = \{m_1, m_2, m_3, m_4\}$. The model features get another index r which represents their sequence: $F = \{m_{1_1}, m_{2_2}, m_{3_3}, m_{4_4}\}$.
To extract the sequence a function pos was introduced:

$$\text{pos}_F : m_{j_r} \mapsto r \in \mathbb{N}^+ \tag{10}$$

results in the index r of a feature m_{j_r} in a model surface F.

As for an interpretation I_n with $n > 1$ the positions of the first associated model feature $I_1.m$ and the last associated model feature $I_n.m$ in the only remaining model surface $F_C \in C_{I_n}$ can be determined:

$$\begin{aligned} \texttt{first} &:= \text{pos}_{F_C}(I_1.m) \\ \texttt{last} &:= \text{pos}_{F_C}(I_n.m) \end{aligned}$$

A new association $(s : m)$ with position

$$\texttt{new} := \text{pos}_{F_C}(m)$$

for model feature m in the context of the previous associations I_n is allowed iff new lies *not* between the indices first and last:

$$\texttt{new} \notin [\texttt{first}, \texttt{last}]_{F_C}$$

with an interval $[\texttt{a}, \texttt{b}]_F$ (with $1 \leq \texttt{a}, \texttt{b} \in \mathbb{N}^+ \leq \text{Card}(F)$ and F a model surface) defined as follows:

$$[\texttt{a}, \texttt{b}]_F = \begin{cases} \{z \in M | \texttt{a} \leq z \leq \texttt{b}\} & \text{, if } \texttt{a} \leq \texttt{b} \\ M \setminus \{z \in \mathbb{N}^+ | \texttt{b} < z < \texttt{a}\} & \text{, else} \end{cases} \tag{11}$$

with $M := \{z \in \mathbb{N}^+ | 1 \leq z \leq \text{Card}(F)\}$.

The claim for the ordering of model features in associations of $\mathcal{C}^3$ can thus be formulated as follows:

$$\mathcal{C}^{3b}_{I_n}((s : m)) \Leftrightarrow (n < 2) \vee (\text{pos}_{F_C}(m) \notin [\text{pos}_{F_C}(I_1.m), \text{pos}_{F_C}(I_n.m)]_{F_C}) \tag{12}$$

with F_C as the one and only remaining model surface in context C_{I_n}.

After a formalization of both aspects of $\mathcal{C}^3$ the constraint can now completely be formulated:

$$\mathcal{C}^3_{I_n}((s : m)) = \mathcal{C}^{3a}_{I_n}((s : m)) \wedge \mathcal{C}^{3b}_{I_n}((s : m)). \tag{13}$$

4 Results

We examine the object recognition results in terms of three values. First, the *number of nodes* in the interpretation tree. A lower number of nodes in the tree means lower expense of the search. Second, the *number of interpretations* generated by the search. This value is used as an intermediate result. And finally the *number of hypotheses* which holds for the most important value. The lower the number of the final hypotheses the lower the extent which is necessary to verify global consistency.
In table 1 we present the recognition results for the example shown in figure 1. Obviously the number of nodes as well as the number of interpretations and especially the number of hypotheses were significantly reduced when using $\mathcal{C}^3$.

	with $\mathcal{C}^3$	without $\mathcal{C}^3$
number of nodes:	877	1336
number of interpretations:	122	213
number of hypotheses:	4	16

Tabelle 1. Results of the recognition process

We tested the system against ten different scenes, each showing one of four different objects. For each test case we gathered the recognition results using and not using $\mathcal{C}^3$. The number of hypotheses occurring when *not* using $\mathcal{C}^3$ is referred to as 100%. The number of hypotheses changes when applying $\mathcal{C}^3$. The relative difference is shown in figure 4. The diagram shows that the costs of computing may be reduced down to an average of 35% of the costs incurred when not using context and ordering constraints.

5 Conclusion

In this work we presented a knowledge-based 3D object recognition approach. The used features are enriched by contextual attributes which we extracted from segmentation results. We showed how an interpretation tree search algorithm can be extended by context and ordering constraints. By applying these constraints we achieve more efficiency in recognition results owing to generating less hypotheses. The use of new introduced constraint $\mathcal{C}^3$ prevents bad hypotheses, e.g. hypotheses containing mirror-inverted associations which often occur because of the local impact of constraints.

So we conclude that the usage of context information in terms of features enriched by attributes leads to more accurate and more efficient recognition results.

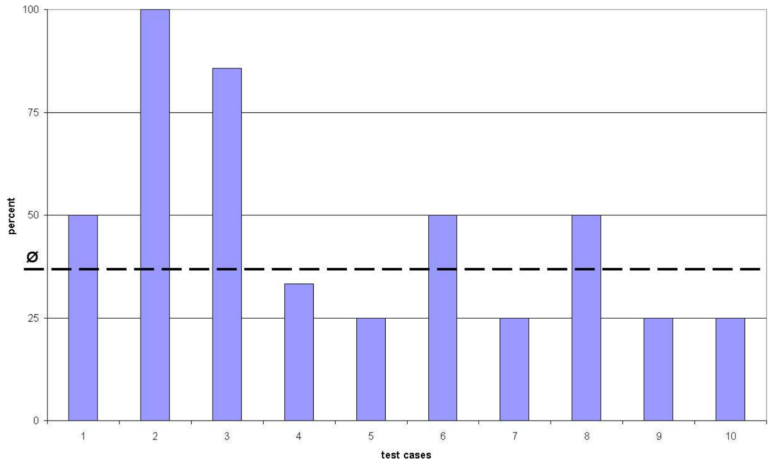

Abbildung 4. Effects of the $\mathcal{C}^3$ regarding the number of hypotheses

Literatur

[AA93] ARMAN, FARSHID and J.K. AGGARWAL: *CAD-Based Vision: Object Recognition in Cluttered Range Images Using Recognition Strategies.* CVGIP: Image Understanding, 58(1):33–48, July 1993.

[BJ01] BASRI, RONEN and DAVID W. JACOBS: *Projective Alignment with Regions.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(5):519–527, May 2001.

[FJ91] FLYNN, PATRICK J. and ANIL K. JAIN: *BONSAI: 3-D Object Recognition Using Constrained Search.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(10):1066–1075, October 1991.

[FWH01] FÖDISCH, MIKE, ANDREAS WINTER und THORSTEN HERMES: *Modellbasierte Raumlageerkennung von Objekten auf Tiefenbildern.* In: *3D-NordOst 2001, Tagungsband: 4. Anwendungsbezogener Workshop zur Erfassung, Verarbeitung, Modellierung und Auswertung von 3D-Daten.* GFaI, Berlin, Dezember 2001.

[Gri90] GRIMSON, W.E.L.: *Object Recognition by Computer: The Role of Geometric Constraints.* The MIT Press, 1990.

[HN00] HORNEGGER, JOACHIM and HEINRICH NIEMANN: *Probabilistic Modeling and Recognition of 3-d Objects.* International Journal of Computer Vision, 39(3):229–251, 2000.

[JHG99] JÄHNE, B., H. HAUSSECKER, and P. GEISSLER (editors): *Handbook of Computer Vision and Appication*, volume 2. Academic Press, 1999.

[PL90] PAVLIDIS, T. and Y. LIOW: *Integrating Region Growing and Edge Detection.* IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-12(3):225–233, 1990.

[SD99] SIM, H.C. and R.I. DAMPER: *A Neural Network Approach to Planar-Object Recognition in 3D Space.* Pattern Analysis & Applications, 2:143–163, 1999.

[SM92] STEIN, FRIDTJOF and GÉRARD MEDIONI: *Structural Indexing: Efficient 3D Object Recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2):125–145, February 1992.