Object Recognition with Shape Prototypes in a 3D Construction Scenario

Martin Hoffhenke and Ipke Wachsmuth

Universität Bielefeld
Technische Fakultät, Arbeitsgruppe Wissensbasierte Systeme
D-33594 Bielefeld
email: {martinh, ipke}@TechFak.Uni-Bielefeld.DE

Abstract. This paper is concerned with representations which enable a technical agent to recognize objects and aggregates from mechanical parts as they evolve in an ongoing construction task. The general goal is that the technical agent has a detailed understanding of the task situation such that it can execute instructions issued by a human user in a dynamically changing situation. A levelled approach for comprehensive shape representation is presented which is motivated by a cognitive model of pictorial shape representations (prototypes). In particular, our system is able to derive and represent spatial properties (such as orientation) and geometric features (e.g., axes or planes) that can be ascribed to the developing construct.

1 Introduction

The context of this work is a human-machine-integrated construction scenario. This paper is concerned with representations which enable a technical agent to recognize objects and aggregates from mechanical parts as they evolve in an ongoing construction task. The general goal is that the technical agent has a detailed understanding of the task situation such that it can serve as an intelligent assistant and execute instructions issued by a human user.

What are the general challenges that have to be met by a representation system employed by the agent? Firstly, the system needs to provide knowledge about the mechanical objects, their properties, and how they can be mounted, and also knowledge about the assembly groups that are being built up in a dynamically changing situation. As is typical in a construction scenario, aggregates of generic building parts form new meaningful objects in which the individual parts can change with respect to their function and role. Secondly, the system should be able to derive and represent spatial properties (such as orientation) and geometric features (e.g., axes or planes) that can be ascribed to the developing construct. A representation of shape could provide a more profound understanding of aggregate objects through enriching the representation by spatial attributes like intrinsic orientations, axis positions or directions (Fig. 1).
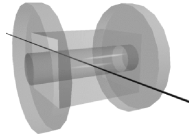
Fig. 1. General shape of an undercarriage. The line is the intrinsic axis which denotes the possible direction of movement restricted by the wheels. An intrinsic orientation (front and back) will be assigned to the undercarriage when mounted to a vehicle.

While the first aspect has been dealt with by way of propositional descriptions in some approaches (e.g., [11], [16]), the problem of representing generic shape has insufficiently been tackled in previous literature. For instance, in earlier work (e.g., [14], [4]), two- or three-dimensional shape models were employed in the recognition and classification of 2D camera percepts. But with respect to construction scenarios in which 3D camera percepts are provided, these approaches are still lacking solutions of the following problems:

- the representation of shape on different levels of abstraction to describe spatial structure at various levels of detail
- the ascription of additional spatial properties by which spatial references could be handled, as in the instruction 'Mount the bumper to the back side of the car'
- a greater complexity of recognition implied by the usage of 3D percepts

The present paper describes an approach to solve these problems, based on the idea of *levelled shape prototypes*. Our scenario is a complex construction task in which a human instructor supervises a robotic agent in the execution of various construction steps. The robot has stereo cameras and is thus able to utilize 3D percepts of the construction scene[1]. Our results have so far been evaluated in a 3D assembly simulation, the Virtual Constructor, that was also used by (Jung et al). The Virtual Constructor can perform all manipulations with virtual building blocks that are physically possible with their real counterparts.

The paper is organized as follows. In Section 2, we give a brief description of the task setting and how the Virtual Constructor is able to simulate construction steps based on a structural (propositional) representation. Section 3 describes the motivations and general goals and features for our novel approach of levelled shape prototypes. In Section 4, we present our approach in more detail and describe results how shape prototypes were used to build up spatial representations in the construction scenario. In Section 5 we conclude that the levelled approach has the advantage to enable both a very abstract shape recognition and an appropriate shape refinement by an integrated representation.

---

[1] To avoid misunderstandings: the percept is a threedimensional reconstruction of the current scene, not a bitplane.

2 A virtual construction scenario

The testbed we use for the evaluation of our shape prototypes is the Virtual Constructor, a tool able to simulate a construction scenario with a human being as instructor and a technical system as constructor. The instructor can manipulate the virtual scene containing a variety of different mechanical objects of a construction kit. Manipulations can be issued by simple natural language instructions (Fig. 2) or directly via the mouse input device. Scene manipulations can be: moving objects, connecting objects, disconnecting objects, and rotating objects within an aggregate.
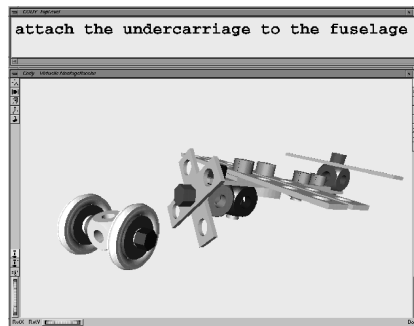

Fig. 2. Screenshot of the Virtual Constructor. The undercarriage will be attached to the airplane via a natural language instruction.

The Virtual Constructor uses a frame language to represent knowledge about the mechanical parts (how they can be connected with other parts) and knowledge about the aggregates which can be constructed. It allows to define a precise building-up of assembly groups by using containment, connection and simple spatial relations like parallelism of two bars. It also allows to define roles that specific building blocks can take on in the context of an assembly group. A bolt for example can become an axle in the context of an undercarriage. This way the bolt can be referred to in the following instructions *by role* like: 'Attach the axle to the green block'. Such a structural description of an assembly group is not very flexible against arbitrary but shape-invariant variations and it is bound to and restricted by a certain construction kit.

Shape prototypes as introduced in the following section are not for defining precise assembly groups out of a special construction kit. Rather, they enrich the construction system with knowledge about the rough appearance of objects belonging to a category. Shape prototypes describe the generic shape of complex objects with parametric geometry models. They are able to represent knowledge about spatial properties of objects like an intrinsic orientation, thus it is possible to handle instructions like: 'Attach the propeller to the front of the airplane'.

3 Object representation with shape prototypes

The idea we develop in this paper is based on a cognitive motivation. A variety of psychological investigations in the last 20 years have shown that human beings store and process information not only in an abstract-structural manner but also often make use of a pictorial concept of things and situations (e.g., [12], [2], [8]). Especially for the description of object categories, prototypical pictorial representations have been postulated. Pictorial representations seem advantageous in that they provide a more 'natural', i.e. intrinsic representation ([13]) of geometrical and spatial properties than propositional representations. It is assumed that by using intrinsic representations the classification of objects should be faster and more robust against variations and should give clues for a subsequent structural analysis, which is a motivation for our technical approach.
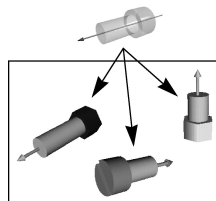

Fig. 3. Property inheritance from shape prototypes: the individual bolts inherit the fitting direction from the bolt shape prototype shown in the upper part.

3.1 Two aspects of shape prototypes

For an intuitive interaction with a human partner the technical agent needs to have a similar understanding of the common subject of their interaction. In the construction scenario aggregates often become meaningful for a human being by their overall shape. For example, the instructor calls two wheels an undercarriage when fitted together by an axle. This is the case when something *looks like* something known. To enable this *act of recognition* in a technical system it is necessary to model the prototypical shape properties of such objects. With structural representations (like frames) it is not possible to recognize objects which are similar by their rough appearance but which are built up in a varied way or with parts of some other construction kit. This is especially true for the recognition of the overall shape of multi-part objects. In contrast, shape prototypes just represent the rough shape and do not take the specific building parts into consideration.

Besides the aspect of recognition there is another important aspect to prototypes when used in the interaction with human beings, which is the *ascription*

of properties even when they are not relevant for classification. But when for example speaking about the front side of an airplane this implies an intrinsic orientation (this means a spatial segmentation by the object's point of view). Another example is shown in Fig. 3. Each individual bolt has a fitting direction that can be inherited from the prototypical bolt shown in the top of the figure. Like classical property inheritance in propositional representations, our aim is a property inheritance of spatial attributes from a pictorial representation. Ascribing such spatial attributes to a part or a construct should enable a technical agent to handle instructions like:

– *Place the bolt upside down.*
– *Attach the propeller at the front side of the airplane.*
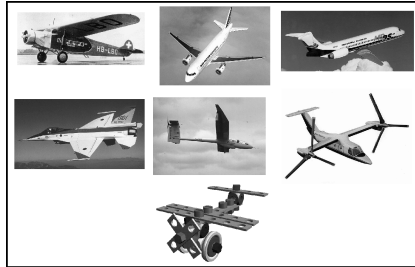– *Show the bottom side of the airplane.*

Fig. 4. Examples of differently shaped entities of the category airplane.

3.2   Levels of abstraction

There are two conflicting requirements to be achieved by a pictorial shape representation. On the one hand it must be able to represent shape in the most abstract manner, in order to recognize a wide variety of objects that have a similar shape. On the other hand it must be able to distinguish between super- and subcategories by discriminating further shape properties. This means that shape prototypes must be able to describe categories at different levels of abstraction, like 'airplane' and 'jet'.

There are a variety of approaches to represent shape found in the literature. Some contributions just deal with 2D representations. For instance [9] uses a boundary representation; [3], [14], [15] are working with symmetry-axis representations. Some early contributions have suggested to use 3D representations ([10], [2]). Especially in image recognition (e.g., [4], [5]), 3D representations are often used to permit a viewer-independent description of objects. But these approaches are restricted to an isolated kind of representation and thus do not permit the modelling of category-adequate abstractions for shape prototypes.

Basically, the above-mentioned approaches restrict to a single type of object or even a single object. They do not provide a universal model of a shape prototype which realizes the conflicting requirements mentioned above. Thus we pursue an integrated approach of gradually abstracted levelled shape prototypes.

The *first requirement* for a shape prototype is to represent shape in the most abstract manner. This can be illustrated by the category airplane. The entities of this category are typically comprised by a wide variety of different shapes (Fig. 4). So, what is the characteristic shape of an airplane? Especially the model airplane built up with blocks of a toy construction kit (Fig. 4, bottom) illustrates that not constructional detail, but overall shape is typical for an airplane. Even with holes in the wings and propeller-blades and untypical features on the airplane's top the object is recognized as an airplane by human beings at first sight. The object is classified as an airplane by its typical spatial layout. Thus the three-dimensional skeleton-model shown in Fig. 5 provides an adequate shape representation for the most abstract shape description of an airplane.
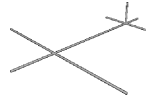

Fig. 5. Shape representation of an abstract airplane (skeleton-model).

The *second requirement* for shape prototypes is the possibility of modelling further discriminating shape properties. A skeleton-model is not sufficient for this because differences often just occur in a two- or three-dimensional expression of a shape (for example, the characteristic delta-shape of the fighter plane in Fig. 4). Thus we need to represent also higher dimensional shape properties. This is why we developed a representation scheme that integrates shape descriptions at levels of different expressiveness.

4   A levelled approach for comprehensive shape representation

As was argued before, we cannot restrict ourselves to a single kind of shape representation, for the requirements on different levels of abstraction are too unlike by far. Thus the aim is to create a framework in which the various shape representations can be integrated.

There are two main aspects to be considered: First, the inclusion of the prototypes in a uniform conception to reach an inter-operability between them and, second, to save the individual properties of the different kinds of shape prototypes in their specific expressiveness. Note that the abstract expressiveness of a prototype is not only relevant for its (geometric) shape-model but also in the classification algorithm working on it.
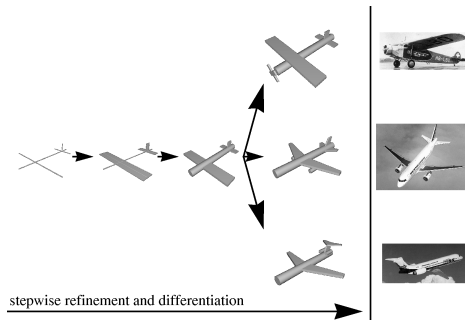
Fig. 6. Cascaded levels of shape prototypes: skeleton-model, plane-structure-model, volumetric model.

To keep the individual characteristics of each kind of prototype, they are embedded (with shape-model and classification algorithm) in a levelled architecture with a uniform input/output-structure and ordered like a cascade with respect to decreasing abstraction, as motivated by Fig. 6. This way the architecture takes care of an increasing differentiation of shape. In our approach we use three different levels of abstraction:

- The *skeleton model* represents objects on the most abstract level. Objects are described just by their major spatial extensions. All three objects in Fig. 7 can be recognized as an airplane on this level. Even the airplane with the swept back wings will be recognized, because on this level of abstraction the 'wing-line' does not stand for a straight line, but for a spatial extent at this position and direction.
- The second level (*plane-structure model*) takes also two-dimensional aspects (flat-shape elements) into account. This is why the left object in Fig. 7 would not be recognized as an airplane on this level (that is, it is not 'as good' an airplane as the other two objects).
- The third level (*volumetric model*) uses three-dimensional shape prototypes. Thus it is possible to discriminate shape elements of the objects by their volumetric shape. The middle and the right airplane in Fig. 7 can be distinguished between more special types of airplane like visualized in Fig. 6.

Any shape model on a more abstract level subsumes all less abstract models.

Fig. 7. Some examples of aggregates recognized as airplanes.

The classification process starts by comparing the percept with the shape prototype on the most abstract level (e.g., the skeleton airplane as represented in Fig. 9). If the shape properties of the prototype matches the percept (like visualized in Fig. 8), a hypothesis (e.g., Fig. 10) is generated. This hypothesis serves as additional information for the matching process with the prototype on the next lower level of abstraction. If it is possible to verify the hypothesis on this level, a new hypothesis with enriched information is generated and passed on to the next lower level.
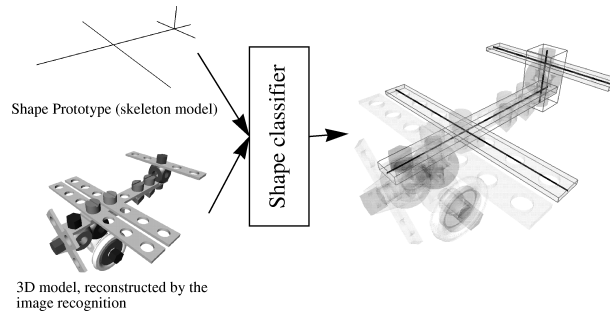
Shape Prototype (skeleton model)

Shape classifier

3D model, reconstructed by the image recognition

Fig. 8. Classification of the model airplane with the shape prototype of an airplane.

4.1 Object recognition with shape prototypes

We now describe in detail how the fully implemented shape recognizer for the skeleton model is used in object recognition. As an example Fig. 9 shows the skeleton prototype of the airplane that is visualized in Fig. 8 (upper left). The prototype contains several shape elements represented by line segments. Each line segment has a default position and some degrees of freedom (tolerance) for its relative position, rotation, and size (`TRANSLATION`, `ROTATION` and `SCALE`). Another important slot is the `Significance`-value. This value rates the expressiveness of a shape prototype as illustrated in Fig. 11. The level of abstraction of an expressive prototype is much lower for a low-structured object (like the wing) than for a high-structured object (like the airplanes in Fig. 6). So the skeleton model of the wing has a much smaller significance value than the airplane has on this level of abstraction.

The recognition process like visualized in Fig. 8 works as follows:

1. Transformation of the main axis of the prototype parallel to the main axis of the object by a principle-component-analysis. For a reduction of ambiguity we make use of the prototype's intrinsic top orientation as a clue for a preferred orientation.
2. For a first rough hypothesis the prototype is scaled to an object-similar size and centered.

```
Modelname: S_AIRPLANE
AbstractionLevel: Skeleton
Significance: 0.8
ShapeElements:
  (fuselage-1 FUSELAGE)
  (wing-1 WING)
  (elevatorunit-1 ELEVATORUNIT)
  (rudderassembly-1 RUDDERASSEMBLY)
LineSegments:
{
  fuselage-1 (CENTER (0, 0, 0) (1, 0, 0) 100) [
    FIX
  ]
  wing-1 (CENTER (10, 0, 0) (0, 0, 1) 90) [
    JOIN (fuselage-1
      STICK
      TRANSLATION ((+/-15unit) (0) (+/-15unit))
      ROTATION (() () ())
      SCALE (+/-15%)
    )
  ]
  rudderassembly-1 (START (-35, 0, 0) (0, 1, 0) 15) [
    JOIN (fuselage-1
      STICK
      TRANSLATION ((+/-10unit) (-5unit) (+/-1unit))
      ROTATION (() () ())
      SCALE (+/-50%)
    )
  ]
  elevatorunit-1 (CENTER (-35, 0, 0) (0, 0, 1) 35) [
    JOIN (rudderassembly-1
      STICK
      TRANSLATION ((+15unit) (0) (+/-10unit))
      ROTATION (() () ())
      SCALE (-50% +130%)
    )
  ]
}
Orientation:
{
  front (1, 0, 0)
  top (0, 1, 0)
}
```

Fig. 9. Coded description of the skeleton-model of an airplane visualized in Fig. 8, left-upper side (range values in TRANSLATION denote tolerances).

3. Search for shape elements of the object (building blocks) which are near and similar-oriented to the main shape element of the prototype (defined as FIX in Fig. 9). The prototype is corrected in position, orientation, and size accordingly.
4. Step by step now the other shape elements are searched for in the object with respect to their degrees of freedom in translation, rotation, and size.

```
Name: AIRPLANE-1
RelatedAggregate: AGGREGATE-9
IP-Type: AIRPLANE
AbstractionLevel: Skeleton
EstimateOfQuality: 0.8
ShapeElements:
{
  ( Name: fuselage-1
    IP-Type: FUSELAGE
    Line: (134.581, 92.3067, 186.292)
      (134.581, 92.3067, -27.7083)
    Parts: H5_BAR-na-2 H3_BAR-na-1 H3_BAR-na-4)
  ( Name: wing-1
    IP-Type: WING
    Line: (26.5817, 98.4755, 48.291)
      (242.586, 98.1379, 48.2924)
    Parts: H7_BAR-na-2)
  ( Name: elevatorunit-1
    IP-Type: ELEVATORUNIT
    Line: (58.0907, 124.427, 172.292)
      (211.095, 124.187, 172.292)
    Parts: H5_BAR-na-1)
  ( Name: rudderassembly-1
    IP-Type: RUDDERASSEMBLY
    Line: (134.621, 76.2946, 172.292)
      (134.621, 138.319, 172.292)
    Parts: BOLT-re-1 BOLT-re-2 BLOCK-re-4)
}
Orientation:
{
  front (0, 0, -1)
  top (0, 1, 0)
}
```

Fig. 10. Description of the instance of the airplane shape prototype in Fig. 8 (right).

If the shape prototype matches the perceived object, an instance of the prototype is created as exemplified in Fig. 10 (and visualized in Fig. 8, right). The instance contains information about the individual shape elements which are passed as clue to the recognizer on the next abstraction level, together with an estimate of quality. Such a clue consists of:

— the type of the object
— the position, orientation and size of the object and its individual shape elements

– the building blocks by which the individual shape elements are formed

This estimate of quality is composed by different factors like the activation of the related category and the significance value of the actual shape-prototype.

4.2 Property inheritance of spatial attributes

As described in Section 3.1, the technical agent needs knowledge about spatial properties that goes beyond the shape aspects important for recognition. So shape prototypes have to contain information that provides additional spatial knowledge about the prototype. As an example, Fig. 9 shows the coded skeleton shape prototype of an airplane. The prototype contains, beside the shape elements, spatial properties describing the intrinsic orientation of an airplane. In this example, the `Orientation`-slots define the front and the top direction. When an instance of a shape prototype is created, it inherits the spatial properties like the intrinsic orientation in Fig. 10 or the fitting direction in Fig. 3.

5 Conclusion and future work

In this paper we presented an integrated approach to model shape prototypes on different levels of abstraction. The object recognizer working on these prototypes is able to recognize and classify objects that are similar to a very general shape prototype on the one hand and, on the other hand, to differentiate among more special prototypes. The shape prototypes can be enriched with spatial attributes that are inherited to the instances of the prototypes. Thus it is possible to ascribe spatial properties like an intrinsic orientation to an object.
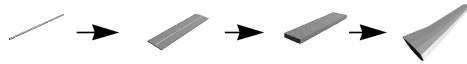

Fig. 11. An expressive prototype for a low-structured object like the wing is provided on a low level of abstraction.

First experiences show that our recognition system based on shape prototypes is up to two times faster than the frame-based recognizer provided with the Virtual Constructor (cf. Section 2). More importantly, it can recognize a diverse variety of different instances of a category like the different airplanes shown in Fig. 7. So far, the recognition time depends much on the complexity of the percept. We expect to get better results by preprocessing the percept with perceptual-grouping methods (e.g., [1]).

Another aspect of future work is the integration of a cognitive process model ([7]) which allows us to take category activation into account. This means that if, for example, some parts or aggregates are recognized as parts of an airplane, then other (generic) parts in this context will be classified more likely as airplane parts, too.

Acknowledgement

References

1. F. Ackermann, A. Maßmann, S. Posch, and D. Sagerer, G. aand Schlüter. Perceptual grouping of contour segments using markov random fields. *Int. Journal of Pattern Recognition and Image Analysis*, 7 (1):11–17, 1997.
2. I. Biederman. Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2):115–147, 1987.
3. H. Blum and R. N. Nagel. Shape description using weighted symmetric axis features. *Pattern Recognition*, 10:167–180, 1978.
4. R. A. Brooks. *Model-based computer vision*. UMI Research Press, Ann Arbor, Michigan, 1984.
5. J. M. Ferryman, A. D. Worrall, G. D. Sullivan, and K. D. Baker. A generic deformable model for vehicle recognition. In *Proceedings of British Machine Vision Conference*, pages 127–136, University of Birmingham, 1995.
6. B. Jung, M. Hoffhenke, and I. Wachsmuth. Virtual assembly with construction kits. In *Proceedings of the 1998 ASME Design for Engineering Technical Conferences (DECT-DFM '98)*, Sacramento (CA), 1998.
7. K. Kessler and G. Rickheit. Dynamische Konzeptgenerierung in konnektionistischen Netzen: Begriffsklärung, Modellvorstellungen zur Szenenrekonstruktion und experimentelle Ergebnisse. *Kognitionswissenschaft*, 8 (2), 1999.
8. S. M. Kosslyn. *Image and brain: the resolution of the imagery debate*. MIT Press, Cambridge (MA), 1994.
9. M. Leyton. A process-grammar for shape. *Artificial Intelligence*, 34:213–247, 1988.
10. D. Marr and H. K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. In *Proceedings of the Royal Society of London B*, volume 200, pages 269–294, 1978.
11. L. Padgham and P. Lambrix. A framework for part-of hierarchies in terminological logics. In *Principles of Knowledge Representation and Reasoning*, pages 485–496. Morgan Kaufmann, San Francisco (CA), 1994.
12. A. Paivio. *Imagery and verbal processes*. Lawrence Erlbaum Associates, Hillsdale (N.J.), 1979.
13. S. E. Palmer. Fundamental aspects of cognitive representations. In E. Rosch and B. B. Lloyd, editors, *Cognition and Categorization*, pages 259–303. Erlbaum, Hillsdale (NJ), 1978.
14. H. Rom and G. Medioni. Hierarchical decomposition and axial shape description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15 (10):973–981, 1993.
15. K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker. Shock graphs and shape matching. In *Proceedings of the Sixth International Conference on Computer Vision*, Bombay, India, 1998.
16. I. Wachsmuth and B. Jung. Dynamic conceptualization in a mechanical-object assembly environment. *Artificial Intelligence Review*, 10 (3-4):345–368, 1996.