Model-based Recognition of 2D Objects in Perspective Images[1]

S. Wirtz and D. Paulus

University Koblenz, Universit@atsstr. 1, D-56070 Koblenz
e-mail: wirtzstefan@uni-koblenz.de; paulus@uni-koblenz.de

Abstract—We report on a case study showing that using symbolic descriptions to recognize objects in perspective images delivers similar results as heuristic or statistical methods. The knowledge is modeled in TGraphs which are typed, attributed, and ordered directed graphs. We combined the search in the state space with a maximum weight bipartite graph-matching and in consequence we were able reduce the numerous amount of hypotheses. Furthermore we used hash tables to increase the runtime efficiency. As a result we are able to show that model based object recognition using symbolic descriptions is on a competitive basis.

DOI: 10.1134/S105466181102115X

INTRODUCTION

While heuristic approaches are commonly used to solve specific object recognition tasks, model-based approaches grant the opportunity for more general solutions. Further, object detection using symbolic descriptions for the model-based image analysis is still a challenge, although there has been research in this area since many years. The advantage of symbolic descriptions is that no large amount of training data is required and man-made objects can usually be described by symbolic descriptions straight forward.

Nevertheless the disadvantage is that it needs a lot of work to create a model (e.g., of a building). Indeed Google Earth provides lots of building-models which are designed in Geography Markup Language (GML) and Keyhole Markup Language (KML) (see Fig. 1). These models deliver symbolic descriptions of buildings which may give new relevance to model-based image analysis, because plenty of models are available.

In this paper we present a case study showing the recognition of 2D objects in perspective images with a *task-independent* pattern recognition approach in the application domains of domino tiles and poker cards.

In a prospective study we are going to apply these results to the more complex 3D domain of building recognition.

Like [1], we use TGraphs [2] as a "light-weight"-ontology and we also use a task-independent *activity-control*, which manages the application flow of the system.

Nevertheless we extend the activity control to work with a *subgraph partition* and an efficient hashing to increase the performance of the system. Furthermore the activity-control employs the Hungarian method, a fast graph- matching algorithm, whenever it is possible to reduce the assignment task to a maximum weight bipartite graph-matching.

One archetype of such a system is ERNEST [3], which is a pattern recognition system developed in 1980 employing semantic networks in several application domains, like building recognition [4] and speech understanding [3]. We evaluate this approach with the two application domains: Domino tiles and Poker cards. Here we are able to handle perspective images and the resulting distortions.

PROBLEM STATEMENT

The aim of this approach is to create an image analysis system, which works independent of the application domain and is able to detect objects, given the symbolic description in a model. These models contain besides the declarative also procedural knowledge, where the procedural knowledge provides the functions to detect model elements in the image and evaluate the assignment of a model element to an

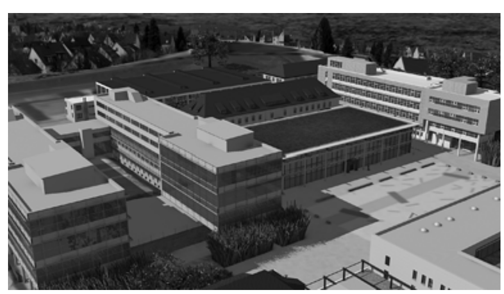

Fig. 1. Google Earth Model of the university Koblenz.

[1]The article is published in the original.

Received January 23, 2011

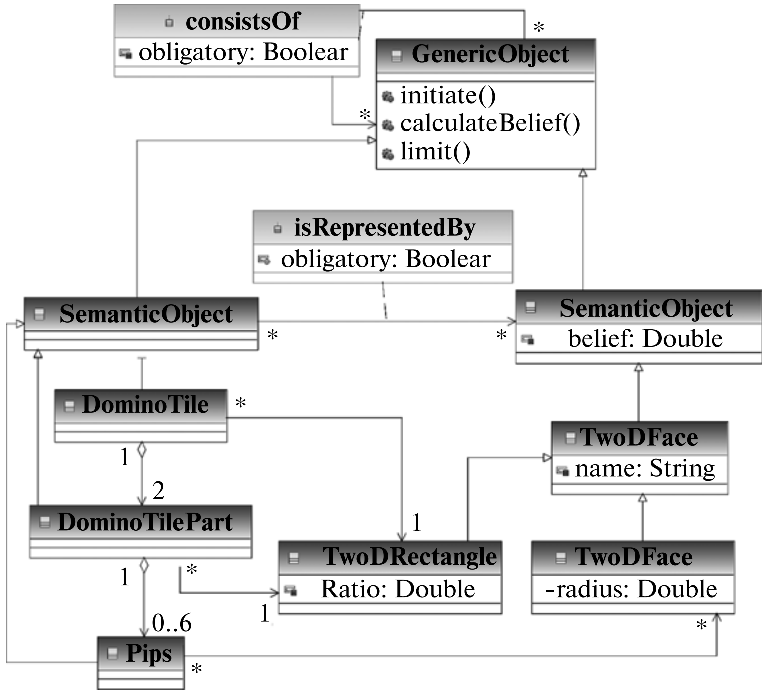

Fig. 2. UML diagram of the domino tile schema.

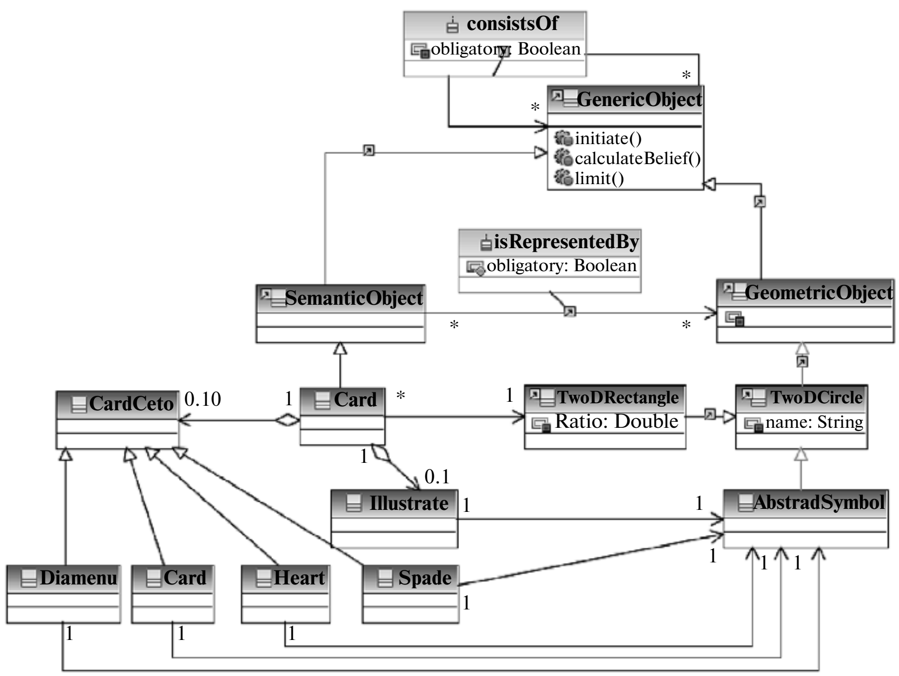

Fig. 3. UML diagram of the poker card schema.

image element. To perform the matching of elements found in the image and elements given by the model, a large number of hypotheses may arise. Therefore, it is necessary to find an efficient way to deal with these hypotheses.

APPROACH AND TECHNIQUES

The basis for the model-based object recognition system is, of course, the model itself. Therefore, we model the declarative knowledge in a generic scheme

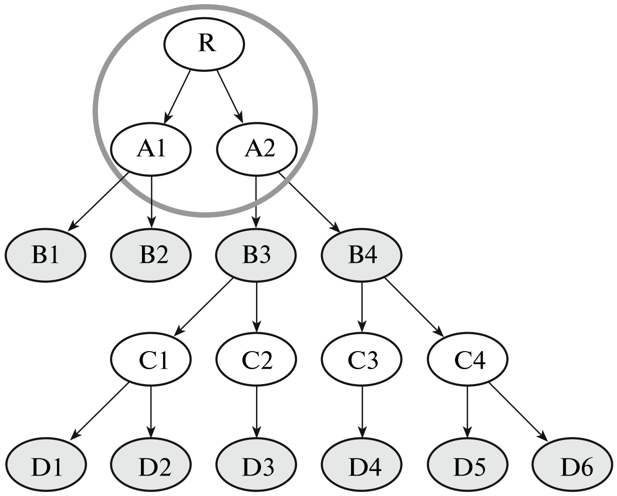

Fig. 4. Model graph of an arbitrary object.

(see Figs. 2 and 3). The main elements of the model are the elements TwoDGeometricObject and SemanticObject, which are derived classes of the GenericObject, so we have a clear separation of semantic and geometric objects.

In addition to the declarative knowledge, we also model procedural knowledge in the schema. The procedural knowledge is available from the *Software techniques for object recognition* (STOR) Component System [5], which supplies an experimental environment for object recognition to construct, analyze and evaluate solutions.

With these schemata, we are able to construct concrete models. In this case study, we have 80 basic models, because we have 28 domino tiles and 52 poker cards.

The approach works similar to [1]. A task-independent and A*-based activity control instantiate the models, fills the state space with hypotheses and controls the application flow.

However, we extended this approach and combined the search in the state space with the Hungarian method. We use this method whenever the cardinality of a node is greater than one.

Therefore, each SemanticObject has to offer a Judgment, according to the belief of Dempster-Shafer theory [6], which describes how well an image element can be assigned to a model element and how missing assignments reduce the belief in the detection of this SemanticObject. These beliefs are combined as described in [1] with the Dempster-Shafer combination rule.

Furthermore, we save all assignments and subgraphs of the models with constant cardinalities in hash tables. This yields to an increase in runtime efficiency, where the detection rate stays equal. We also save the subgraphs with constant cardinalities, because these sub structures are often redundant for the models.

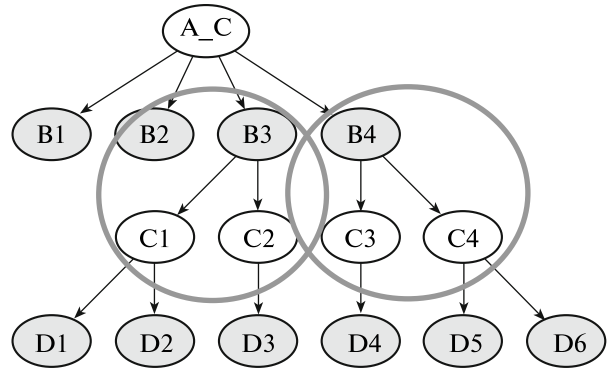

Fig. 5. Reduced model graph.

Fig. 6. Example of a poker card image of the data set.

For instance, Fig. 4 shows a possible model of an arbitrary object. There we assign the elements A1 and A2 with Hungarian method and if the cardinality between the root node R and the node A is constant we will save the whole subgraph in a hash table. As input for the hash function we use the type of model node, the region of interest, the attributes of the node and for subgraphs the types of the edges as well.

Accordingly we exchange the subgraph with a single node, which reduces the complexity of the model. Iteratively we continue the analysis process (see Fig. 5) until all probable models, where the belief of the model exceeds some threshold, are completely assigned or failed to assign.

RESULTS

For the experiments, a data set containing 800 images of domino tiles and poker cards with perspective distortions was used (see Fig. 6).

The play card colors were classified with Hu moments as features, which are invariant under translation, changes in scale, and also rotation [7]. This classification is very robust and delivers good results for the detection.

The overall detection rate of domino tiles and poker cards is about 90% for the dataset. This case study shows that the advantage of the approach is that it handles missing and imprecise data very well and is able to use the segmentation algorithm for specific regions, so that the segmentation results are better than the results of segmentations over the whole image.

CONCLUSIONS

We introduced in a case study an approach which uses symbolic descriptions to recognize 2D objects in perspective images. For this purpose we combined the search in the state space with a graph-matching algorithm to deal with numerous hypotheses. Furthermore we use hashing functions for nodes and subgraphs taking advantage of the redundancy between the models to increase the runtime efficiency.

In a next step we are going to extend this approach to the 3D domain of building recognition.

ACKNOWLEDGMENT

This work was partially funded by Deutsche Forschungsgemeinschaft (DFG) under grant PA 599/8-1.

REFERENCES

1. S. Wirtz, M. Häselich, and P. Dietrich, "Model-Based Recognition of Domino Tiles Using TGraps," in *Proc. 32nd DAGM Pattern Recognition* (Darmstadt, 2010), pp. 101–110.
2. J. Ebert, et al., "Graph Technology in Reverse Engineering, the TGraph Approach," in *Proc. 10th Workshop Software Reengineering (WSR 2008)*, Ed. by R. Gimnich, U. Kaiser, J. Quante, and A. Winter (Bonn, 2008), Vol. 126, pp. 67–81.
3. H. Niemann, et al., "Ernst: A Semantic Network System for Pattern Understanding," IEEE Trans. Pattern Anal. Mach. Intell. **12** (9), 883–905 (1990).
4. F. Quint, "Kartengestützte Interpretation monokularer Luftbilder," PhD Thesis (Univ. Fridericiana zu Karlsruhe, 1997).
5. K. Falkowski and J. Ebert, "The STOR Component System," Tech. Rep. (Institut für Softwaretechnik, univ. Koblenz–Landau, 2009).
6. G. Shafer, *A Mathematical Theory of Evidence* (Princeton Univ. Press, 1976).
7. M. K. Hu, "Visual Pattern Recognition by Moment Invariants," IEEE Trans. Int. Theory **IT-8**, 179–187 (1962).