Learning Qualitative Models from Image Sequences

Jure Žabkar and **Ivan Bratko**
University of Ljubljana,
Faculty of Comp. and Inf. Science,
Tržaska 25,
SI-1000 Ljubljana, Slovenia

Gregor Jerše
University of Ljubljana
Faculty of Mathematics and Physics
Jadranska 19,
SI-1000 Ljubljana, Slovenia

Johann Prankl and **Matthias Schlemmer**
Vienna University of Technology,
Automation and Control Institute,
Gusshausstrasse 27-29 / E376,
A-1040 Vienna, Austria

Abstract

In this paper, we describe the autonomous learning of qualitative models with a robot's on-board vision. Those models are used to describe spatio-temporal qualitative relations between observed objects. Therefore, the algorithm QING is described which extracts the necessary qualitative relations between the objects from the sequence of images. The robot uses these features together with other sensory data to learn about the environment.

Keywords: qualitative (spatial) reasoning, cognitive vision, cognitive robotics

Introduction

In this paper we tackle certain problem from the field of cognitive robotics, namely how the robot can use its on-board vision for autonomous learning. This problem is highly connected to the field of cognitive vision as the robot should somehow reason about the information it gets from image sequences. Following the definition of (Vernon 2008), "A cognitive vision system can achieve the four levels of generic computer vision functionality of detection, localization, recognition, and understanding." Whereas classical computer vision is mainly concerned with the first three points, the last issue affords interdisciplinary work in order to integrate higher-level reasoning functions. This work aims at incorporating a specific machine learning technique in order to qualitatively reason about the arrangement of objects. The abstraction from quantitative pixel data to a more qualitative layer seems to be of great importance to cognitive vision. In this abstraction step, the vision part is concerned with segmenting the image to proto-objects (groupings of pixels that are likely to belong to the same object). In this paper, we are mainly concerned with the learning part, therefore the proto-object grouping is assumed to be given. However, higher-level qualitative reasoning is highly relevant for providing feedback to the vision part, as its ability to predict the existence and the arrangement of proto-objects in the subsequent image(s) can support low-level image processing.

From a roboticist's perspective, a qualitative model at any layer can help interpreting a given situation. This paper tries to bridge this link for one of the lowest, namely the perceptual layer, providing a model for visual scene interpretation. Motivation for this work comes from the European project XPERO, in which a robot should gain insights about the real world by experimenting and meaningfully relate it's intero- and exteroceptive information so to arrive at a level of qualitatively understanding its environment.

The main idea of this paper is to apply the algorithm QING (see corresponding Section) to extract qualitative spatio-temporal features from an image sequence and use them together with other sensory data for autonomous learning about the concept of occlusion. In this paper we present an artificial scenario in which the robot circles around two balls of different colour and builds a qualitative model in the form of a qualitative non-deterministic finite automaton (qNFA). The robot learns autonomously without any external intervention. The final model enables us as well as the robot to reason about the occlusion, e.g. it tells us that it is not possible to go from the state of non-occlusion directly to the state of total occlusion but rather through the state of partial occlusion. Our basic goal is to build a system which the robot could use for reasoning about its visual input and based on this reasoning improve its visual perception, e.g., detecting regions of interest. Our system can discover qualitative relations between the objects in the images and how these relations change over time. Currently, it is capable of discovering topological relations.

As this paper focuses on the use of QING for learning the relations, we will not describe the vision part. We must mention though that the extraction of the "interesting" colour blobs from the images can be motivated by a simple curiosity mechanism. For a robot tuned to learn about sensory input it has not seen before, colour is a strong cue. More elaborate techniques, such as the bottom-up Grouping of line features, as described in the next Section, can be applied as well.

Algorithm QING

QING (Žabkar *et al.* 2007) is an algorithm for qualitative analysis of continuous class variable f w.r.t. given attributes $(x_1, \ldots, x_n)$, where n is the dimension of the attribute space. For simplicity we will in this short descrip-

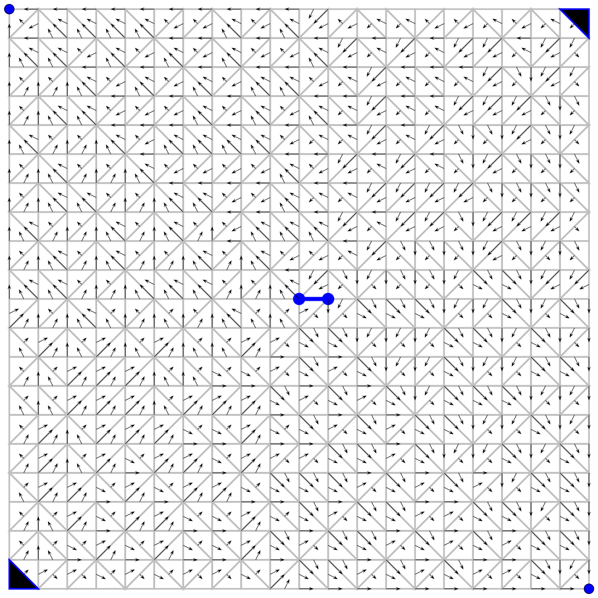

Figure 1: Qualitative field for $f(x, y) = xy$. The arrows point in the direction of function decrease.

tion restrict ourselves to two attributes. Theoretically, QING works for any dimension n, but is practical for $n \leq 5$ due to the complexity of triangulation.

To be more illustrative, we accompany the description of the algorithm with a simple example $f(x, y) = xy$ defined on an orthogonal mesh (see Fig. 1) on the domain $[-10, 10] \times [10, 10]$. Learning examples are represented as points in the attribute space, each point having assigned a value of its class variable. The domain is triangulated in order to be analysed with discrete Morse theory. Critical points, i.e. maxima, minima and saddles, are reconstructed using the algorithm of (King, Knudson, & Mramor Kosta 2005). The output of QING is a qualitative field (Fig. 1), a set of critical points and a labeled qualitative graph (Fig. 2), which is a visualisation of the qualitative model. Detailed definitions of these terms are given in (Žabkar *et al.* 2007). The main difference between QING and other algorithms for induction of qualitative models is in attribute space partitioning. Unlike algorithms that split on attribute values (e.g. trees, rules), QING triangulates the space (domain) and constructs a qualitative field which for every learning example tells the directions of increasing/decreasing class.

The example image, in the experiment that we describe in the next Section, is processed in a similar way. However, to capture the time, we need to connect the neighboring images in the sequence. To do this, we use parametric Morse theory with time as a parameter and we follow the critical simplices through the slices as described in (King, Knudson, & Mramor Kosta 2007).

Experiments

We performed the experiments on artificial data in a domain in which the robot circles around a red and a blue ball as shown in Fig. 3. The robot uses an overview camera to measure the distances to the balls ($bdred$ and $bdblue$) and it uses an on-board camera to observe the balls and collect the data for learning a qualitative model.

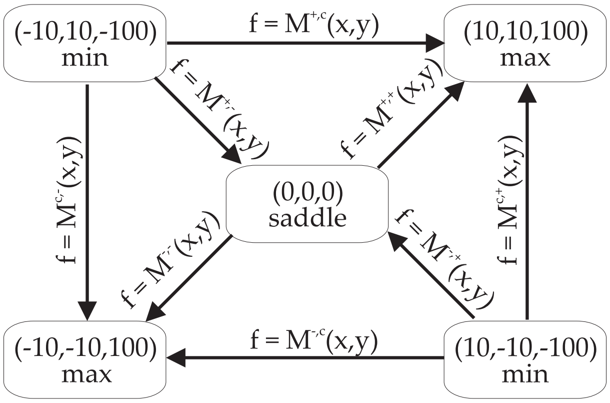

Figure 2: Qualitative graph for $f(x, y) = xy$.

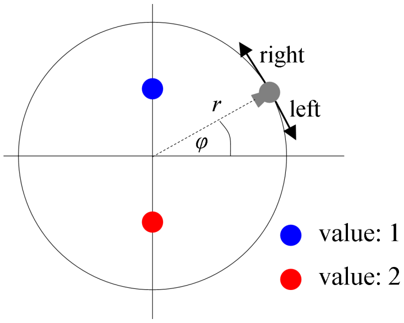

Figure 3: The robot circles around the balls and observes them with its on-board camera.

The robot is also aware of its polar coordinates, so it knows about its angle and the radius. Its actions are *left* (clockwise) and *right* (counter clockwise along the circle as shown in Fig. 3). The robot observes the qualitative change of its distances to the red and the blue ball w.r.t. the action. For example, if the robot resides at $(\varphi = 0\,°, r)$ and chooses to go right, i.e. φ increases, the distance to the red ball would increase while the distance to the blue ball would decrease, $bdred = Q(+\varphi)$ and $bdblue = Q(-\varphi)$. The robot can observe similar relations in the image sequence. The relations that it can detect on a simple image of two balls are the following (see also Fig. 4):

- the red ball and the blue ball do not touch
- the balls touch
- only the red ball is visible
- only the blue ball is visible

Inside QING algorithm, the objects are distinguished by their colour and each colour is represented by a unique numerical value. In our example, we define that the value 1

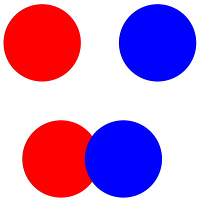

Figure 4: The images from the on-board camera where the balls don't touch (top) and when they overlap (bottom).

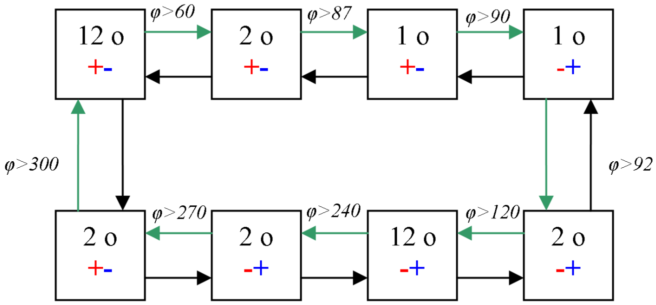

Figure 5: The learned qualitative NFA describing how the robot can change its states with the chosen actions (right...green arrows; left...black arrows). The red signs stand for $bdred = Q(s\varphi)$ and the blue signs are $bdblue = Q(s\varphi)$.

stands for the blue color and the value 2 stands for the red one. The background has value 0. QING constructs a discrete vector field in a 3-dimensional space (2D image and time) and assigns the appropriate values to each pixel, according to the color of the pixel. Although QING can handle noise well, there was no noise in our artificial data. To obtain the qualitative relations QING computes extreme points in this space and uses the discrete vector field to track the movement of these extrema over time.

When the balls don't touch, QING finds two maxima with values 1 and 2 (we mark such a state with '12'). If only one of the balls is visible, it finds either 1 (for blue) or 2 (for red), while if they touch it finds one maximum with value 2. In our image sequence, the changes are rare since most of the time the image at angle φ_i is the same as φ_{i+1}. This, we denote as 'o', meaning there is no change, i.e. steady. Considering the type of the qualitative relation in each image and $bdred = Q(+\varphi)$ and $bdblue = Q(-\varphi)$ from above we build a class value for each learning example, e.g. 12o +-, meaning that the balls stay separated, the distance to the red ball increases (the first sign) and the distance to the blue ball decreases (the second sign).

Results

The learned model is in the form of qualitative nondeterministic finite automaton (NFA) as shown in Fig. 5. Its states are qualitative descriptions of the observations derived from the image features and other available attributes, i.e. distances to both balls and the angle φ. The transitions explain the possible changes of states given an action. Nondeterminism is hidden in the fact that the same action applied in the same state may result in the same state or the neighbouring one, i.e. self transitions are always possible. This is due to the qualitative descriptions of the states. However, such NFA gives us enough information to reason qualitatively about the system. We can observe that total occlusions ($\varphi = 90$ or $\varphi = 270$, changes of $+, -$ signs) may only happen from partial occlusions (states with *2o*).

Related work

Many authors have addressed the problem of qualitative spatial or spatio-temporal reasoning. (Cui, Cohn, & Randell 1992) describes an envisionment-based qualitative simulation program that can reason about space and time, considering the topological relations between objects. Learning temporal patterns from unannotated video data is presented in (Fleischman, Decamp, & Roy 2006). Well known theoretical approaches to qualitative spatio-temporal reasoning are described in (Cohn & Hazarika 2001) and (Randell, Witkowski, & Shanahan 2001). The latter is especially interesting for us as it considers spatial occlusion. (Cao, Mamoulis, & Cheung 2005) discovers sequential patterns in a spatio-temporal series of movements of mobile objects. An interesting approach to mining temporal patterns in multivariate time series, using Unification-based Temporal Grammars is described in (Mörchen & Ultsch 2004). It only considers the temporal dimension but there seems to be no reason against applying a similar technique to spatial dimensions. On the other hand, (Bailey-Kellogg & Zhao 2004), (Lundell 1994) and (Faltings 1995) study only qualitative spatial reasoning.

Concerning the vision part, literature on computer vision is extremely diversified, wherefore we are focusing here on low-level algorithms powerful enough to support the task at hand as well as State-of-the-Art attempts to fuse qualitative reasoning with computer vision.

For grouping pixels to likely objects (so-called proto-objects), a recent work is (Zillich 2007). In this work, edges are grouped based on Gestalt principles, e.g., continuity and proximity. Using a parameter-free anytime algorithm, this tool is capable of delivering the most likely locations of proto-objects in terms of closures and ellipses very fast. Alternatively, colour-based segmentation can be used, for example the graph-based method of (Felzenszwalb & Huttenlocher 2004).

Work on fusing qualitative reasoning with vision techniques has been done by (Bennett *et al.* 2008). In this paper, the authors recognise and track multiple objects throughout a scene (e.g., basketball players) supported by a reasoning about the spatio-temporal continuity. (Huang & Essa 2005) are tracking multiple objects through complex occlusion situations, where a colour blob tracker is backed by a reasoning step of where currently unseen objects are. Their task is very similar to ours except they are using genetic algorithms to match the objects from one image to the next one while we use parametric discrete Morse theory to do this.

Discussion and future work

The above work shows a promising direction towards an autonomous robot system with on-board vision that could learn from the vision input as well as improve on visual perception using qualitative models. We believe that our approach can help the robot extract dynamic features from its vision system and use them in qualitative models. Using these features the robot can detect the region of interest in its environment. The latter is especially interesting combined with the task of embodied learning by experimentation where regions of interest may drive the robot to interact with the world.

From the technical perspective, our future work will include further improvement of the QING algorithm. We would also like to investigate how the vision part can make use of qualiative models, e.g. to improve the image segmentation.

Acknowledgment

The work described in this article has been funded by the European Commission's Sixth Framework Programme under contract no. 029427 as part of the Specific Targeted Research Project XPERO ("Robotic Learning by Experimentation").

References

Bailey-Kellogg, C., and Zhao, F. 2004. Qualitative spatial reasoning extracting and reasoning with spatial aggregates. *AI Magazine* 24(4):47–60.

Bennett, B.; Magee, D. R.; Cohn, A. G.; and Hogg, D. C. 2008. Enhanced tracking and recognition of moving objects by reasoning about spatio-temporal continuity. *Image Vision Computing* 26(1):67–81.

Cao, H.; Mamoulis, N.; and Cheung, D. W. 2005. Mining frequent spatio-temporal sequential patterns. In *ICDM*, 82–89. IEEE Computer Society.

Cohn, A., and Hazarika, S. 2001. Continuous transitions in mereotopology.

Cui, Z.; Cohn, A. G.; and Randell, D. A. 1992. Qualitative simulation based on a logical formalism of space and time. In *National Conference on Artificial Intelligence*, 679–684.

Faltings, B. 1995. Qualitative spatial reasoning using algebraic topology. In *COSIT*, 17–30.

Felzenszwalb, P., and Huttenlocher, D. 2004. Efficient graph-based image segmentation. *International Journal of Computer Vision* 59(2):167–181.

Fleischman, M.; Decamp, P.; and Roy, D. 2006. Mining temporal patterns of movement for video content classification. In *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, 183–192. New York, NY, USA: ACM.

Huang, Y., and Essa, I. 2005. Tracking multiple objects through occlusions. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, 1051–1058.

King, H. C.; Knudson, K.; and Mramor Kosta, N. 2005. Generating discrete morse functions from point data. *Exp. math.* 14(4):435–444.

King, H. C.; Knudson, K.; and Mramor Kosta, N. 2007. Birth and death in discrete morse theory. *Preprint*.

Lundell, M. 1994. Qualitative reasoning with spatially distributed parameters. In *International Workshop on Qualitative Reasoning about Physical Systems*, 13–20.

Mörchen, F., and Ultsch, A. 2004. Mining hierarchical temporal patterns in multivariate time series. In Biundo, S.; Frühwirth, T. W.; and Palm, G., eds., *KI*, volume 3238 of *Lecture Notes in Computer Science*, 127–140. Springer.

Randell, D. A.; Witkowski, M.; and Shanahan, M. 2001. From images to bodies: Modelling and exploiting spatial occlusion and motion parallax. In *IJCAI*, 57–66.

Vernon, D. 2008. Cognitive vision – the development of a discipline. online at: `http://www.eucognition.org/ecvision/about_ecvision/Cognitive_Vision.pdf`.

Žabkar, J.; Jerše, G.; Mramor, N.; and Bratko, I. 2007. Induction of qualitative models using discrete morse theory. In *Proceedings of the 21st Workshop on Qualitative Reasoning*.

Zillich, M. 2007. *Making Sense of Images: Parameter-Free Perceptual Grouping*. Ph.D. Dissertation, Vienna University of Technology.