A Vision prototype for OpenNARS

Robert Wünsche

Technische Universität Dresden
Dresden, Germany
rt09@protonmail.com

Abstract. AGI systems need to operate without interruption. Hence, a vision system for an AGI system has to be active, adaptive and work under the Assumption of Insufficient Knowledge and Resources (AIKR). This paper describes details of a prototype for a domain independent vision system which learns online and incrementally over the lifetime of the agent under AIKR. This prototype is intended to become the basis of a more capable vision system for NARS based AGI systems.

Keywords: Vision · Unsupervised Learning · Object Tracking

1 Introduction

True autonomy can only be reached if all components of the agent, including the artificial brain of the agent, runs without any interruption. This can only be achieved without a seperation between training and operating phase. A vision system for such an agent also needs to be adaptive and work under the Assumption of Insufficient Knowledge and Resources (AIKR, see [1]), and effectively has to work in real time. A prototype for a domain-independent vision system that fulfills that property is proposed. It learns online and incremental over the lifetime of the agent, and operates under AIKR, meaning in real time and with finite processing capacity, while accepting new input at any time. The proposed prototype is intended to become the basis of a more capable vision system for OpenNARS based AGI systems.

2 Methodology

AGI systems require specific modalities which, were designed with the requirements of AGI systems in mind. Requirements include online incremental life-long learning, real-time operation and active vision, all under AIKR. This rules out all contemporary “vision” systems known to the authors, such as YOLOv3 [5], because these systems can’t learn over the lifetime of the agent. Also it is usually impossible to engineer them in such a way that they work as active vision systems [4]. Other issues include the lack of transparency of the AGI in question. A complete vision system must be an active process, working seamlessly together with the AGI it is used in [4]. Ideally perception should also be a transparent process to the AGI system, though it’s not clear to what extent it’s necessary.

Design Requirements The described prototype is designed according to the requirements of vision (pre)processing systems of AGI systems:

– Online Incremental Life-Long Learning [2] - an autonomous system can't rely on pretrained models. It may encounter novel objects and categories at any time.
– Real-Time - the agent can't depend on a "learning phase" in a real-time environment
– AIKR [1] Processing has to happen under finite compute resources because it has to work in real-time with a fixed amount of computer memory and CPU/GPU/etc. compute cycles. When new information is incoming, it can be analyzed and stored with an arbitrary level of detail, so the system has to decide which information to retain and which to throw away (either completely or by compressing non relevant information away). Also the system has to work with insufficient knowledge: new stimulus can arrive at any time and potentially holds information the system did not have so far. While insufficiency of knowledge is a clear requirement for any learning system, insufficiency of knowledge is usually not assumed by others, except of some real-time data mining approaches.

3 Architecture

Preprocessing Preprocessing may be done for better generalization, which is the case for edge filters. This is currently not done in the prototype. Other preprocessing steps may include the detection of optical flow or change in brightness. All of these preprocessing steps are roughly equivalent to what the Retina or V1/V2 are implementing in the human brain.

Reasoning Novel raw sub-symbolic data in the form of pixels has to be analyzed for novel and known regularities. The perception system has to have biases to do so. It was argued in [7] that systems should have more biases than traditional contemporary computer vision state-of-the-art approaches have to date. This is realized in this prototype with a very strong bias towards motion. This makes sense to the authors, because parts of objects tend to move together through the perceived scene. Other biases may be contours, contours may get learned by training a classifier with sub-images of the boundaries of already recognized objects. This makes sense to the authors, because human perception seems to have a weak bias toward this "feature". Another strong bias in human perception is the grouping by color [8]. Contour and color biases are currently not implemented in the prototype. All biases enrich the raw sub-symbolic information and are stored and reasoned about in a symbolic form.

Motion bias Motion between images is tracked with particles which follow the optical flow. The current implementation "seeds" particles in regions with a change between the current frame and the last frame. Sub-images are remembered for each particle to track in future frames. All particles are grouped by motion direction and proximity. This process is repeated for each frame without carrying any information between frames. The grouping is done in a way so that particles which are close together and move with roughly the same velocity are grouped into the same group. Note here that the groups can overlap. An example of an overlap is when a moving object is in front of a moving background. All these groups are stored as "region proposals" for further processing in each frame. Each "region proposal" has a convex boundary in 3 directions (x, y and xy).

Tracking A vision system has to be able to collect information of objects or parts of objects of novel stimuli. This software creates prototypes of tracked objects where a "region proposal" frames a sufficiently novel region in the image. Prototypes are compared by a distance metric. A new Prototype is created if the distance metric is above a threshold (which is provided as a parameter). Prototypes are stored as arrays of pixels where each channel has an estimator for a central valued distribution. This allows to "mask out" pixels which were considered as irrelevant in earlier reasoning processes. Currently pixels which didn't change are considered as irrelevant, so the samples are not stored in the distributions of the channels of the pixels. Tracking of objects is done by searching a closest match of the prototype in a radius around the last known position. Tracking is dropped if the metric of the best match is above a threshold. This treatment was inspired by the predator algorithm [3].

Attention Attention is necessary because of AIKR. Currently only bottom-up attention (controlled by the vision system and sourced by stimuli) is implemented, the system is already flexible enough to implement top-down attention (as controlled by OpenNARS with operators) in the future. Attention is done to obtain the best recognition sampling positions. This is done by sampling an attention map, (super)pixels with higher values get a higher chance to get sampled. The probability of a sampled (super)pixel is equal to the value of the (super)pixel divided by the sum of all (super)pixels. The attention map is currently obtained from the change of the color of a super-pixel. This is remapped by x*x to favor areas with a higher change more strongly.

Recognition The recognition of learned objects is done by classifying sub-images at the positions sampled by the bottom up attention mechanism. Recognition of the learned prototypes is currently not fully implemented and tested. The matched distance may get mapped to a NARS-Truth-Value by:

$$conf = 1.0/(1.0 + dist * distToConfFactor)$$

Where conf is the confidence of the event which will be fed into OpenNARS [6]. The frequency of the truth value is 1.0. distToConfFactor is a parameter which is manually tuned.

4 Discussion and Future Work

Active Vision Active Vision [4] with operator execution is currently not implemented. This is necessary for a AGI system for the reasons which were discussed in earlier sections of this paper. A possible way to implement Active

Vision may be to build action sequences with the context of a cursor. Classifications at the cursor position can be stored as the components of a sequence. The building, comparing and traversal of these sequences can be implemented with specialized processes for the sake of a fast and controlled execution. OpenNARS may be able to query these sequences and modify them if necessary.

Features A Vision System requires more "feature" recognition processes which complement each other for different environments and different lighting conditions. The current prototype can only track objects in relatively simple environments.

Interfacing with OpenNARS An interface to and from OpenNARS is currently missing too, because the authors focused on the required mechanisms of a vision system for an AGI-system. OpenNARS has to be informed about the current perceived situation for each frame. This will be done with events. OpenNARS will be able to actively query the vision system [4]. An example of this interaction is the direction of attention (which is always top-down attention). Another example of this may be the querying of information about a perceived object or relations between objects.

References

1. Wang, P. (1995). Non-axiomatic reasoning system: Exploring the essence of intelligence. Bloomington, IN: Indiana University.
2. Bekel, H., Bax, I., Heidemann, G., & Ritter, H. (2004, August). Adaptive computer vision: Online learning for object recognition. In Joint Pattern Recognition Symposium (pp. 447-454). Springer, Berlin, Heidelberg.
3. Kalal, Z., Matas, J., & Mikolajczyk, K. (2009, September). Online learning of robust object detectors during unstable tracking. In 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops (pp. 1417-1424). IEEE.
4. Wang, P., & Hammer, P. (2018, August). Perception from an AGI Perspective. In International Conference on Artificial General Intelligence (pp. 259-269). Springer, Cham.
5. Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.
6. Hammer, P., Lofthouse, T., Wang, P. (2016, July). The OpenNARS implementation of the non-axiomatic reasoning system. In International conference on artificial general intelligence (pp. 160-170). Springer, Cham.
7. Marcus, G. (2018). Deep learning: A critical appraisal. arXiv preprint arXiv:1801.00631.
8. Lafer-Sousa, R., Conway, B. R., & Kanwisher, N. G. (2016). Color-biased regions of the ventral visual pathway lie between face-and place-selective regions in humans, as in macaques. Journal of Neuroscience, 36(5), 1682-1697.