

# A Prototyping Environment for Integrated Artificial Attention Systems

Jan Tünnermann, Markus Hennig, Michael Silbernagel, and Bärbel Mertsching

GET Lab, University of Paderborn, Pohlweg 47–49, 33098 Paderborn, Germany  
{tuennermann,hennig,silbernagel,mertsching}@get.upb.de  
<http://getwww.upb.de>

**Abstract.** Artificial visual attention systems aim to support technical systems in visual tasks by applying the concepts of selective attention observed in humans and other animals. Such systems are typically evaluated against ground truth obtained from human gaze-data or manually annotated test images. When applied to robotics, the systems are required to be adaptable to the target system. Here, we describe a flexible environment based on a robotic middleware layer allowing the development and testing of attention-guided vision systems. In such a framework, the systems can be tested with input from various sources, different attention algorithms at the core, and diverse subsequent tasks.

## 1 Introduction

The robot operating system (ROS) is a popular robotics middleware. Among other features, it provides hardware abstraction, package management and a communication layer based on the publisher-subscriber paradigm allowing interaction of different components (ROS nodes) [7]. Even though the aim of ROS is simplifying system integration in robotics, it is highly useful in the development and testing of general purpose attention systems. With the concept of messages of certain types, which are published and subscribed by the components, an interface is available that allows a convenient exchange of components.

We propose an architecture consisting of an *input layer* publishing images, which can be obtained from various sources. The *attention layer*, which is the core of the system, performs the selective mechanisms on the images and forwards them to the *task layer* (see fig. 1). Especially during the development phase, an additional *visualization layer* is useful to convert the output at various stages of the system into a form which is comprehensible for the developer. In the following, we describe the layers and the contained components consisting of ROS nodes.

## 2 The Proposed Architecture

The **input layer** performs the image acquisition. Inside the input layer block shown in fig. 1, several examples are listed. Images can be obtained from various

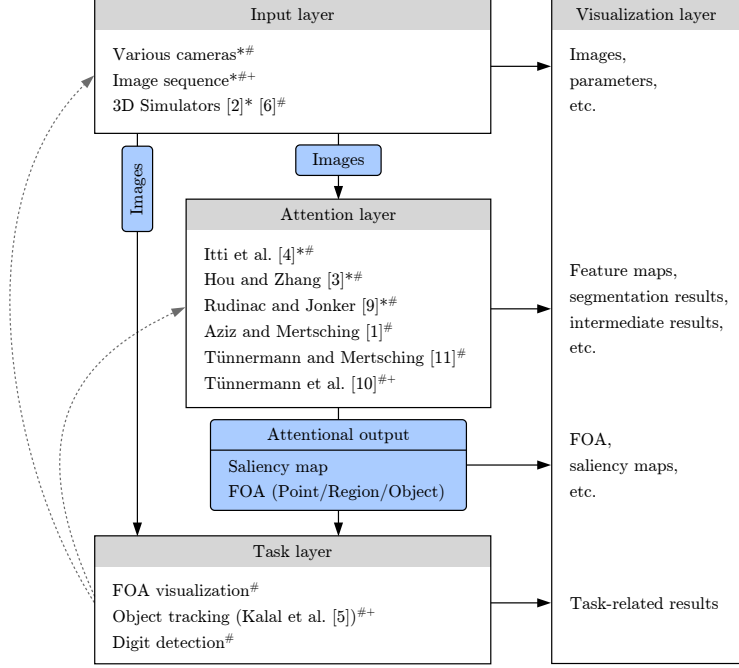


Fig. 1: Proposed ROS-based environment. Besides the different layers processing the data, the main messages, which are subject to synchronization, are shown (blocks with rounded corners; messages for the visualization are not shown). Inside the blocks representing the layers, exemplary instances of components are listed. Publicly available components (as a part of ROS or public ROS packages) are marked with “\*”. Those marked with “#” have been explicitly used in such an environment in our lab. ROS implementations of the models described in [4], [3] and [9] were obtained from the ROS repository [8] and were adapted to allow for synchronization of input and results. Nodes marked with “+” are part of the example described in section 3. The dashed gray arrows represent feedback loops which are subject of future investigations.

cameras (such as webcams, pan-tilt-zoom network cameras, or cameras specialized for computer vision) or fed into the system from recorded images or videos. For many cameras, drivers exist in the form of ROS packages. Tools for feeding images or videos into the ROS system are also available. Furthermore, 3D simulators such as SIMORE [6] and Gazebo [2] have been adapted to publish images in ROS. Additional to image acquisition, preprocessings, such as smoothing or resizing the images, are possible parts of the input layer.

The **attention layer** processes the images by applying attention algorithms (see fig. 1 for a list of algorithms which have been tested as ROS nodes). Some algorithms, such as the spatiotemporal approach described in [11], may require the collection several images to form a spatiotemporal context. Different algorithms provide different data at the output end: While most approaches provide saliency maps and a punctual focus of attention (FOA), others additionally con-

tain areal information such as region outlines or bounding boxes. The latter are considered as “Region-FOA” and “Object-FOA” in fig. 1, indicating that they can cover partial objects (regions) or complete objects.

The **task layer** contains post-attentional tasks. Examples are object tracking or letter detection components, which are initialized with data from the attentional system. Depending on the task, the attentional output, supplementary processing, such as thresholding saliency maps to obtain bounding boxes, may be required.

The **visualization layer** can contain standard tools of ROS (e.g. the image viewer) or custom components (e.g. the volume viewer; see fig. 2 ②) that receive input from other layers.

### 3 Example Instances

Fig. 2 depicts an example instance of the proposed environment, which is used to test a prototypical integration of the system proposed in [10] with an object tracker. This *Tracking-Learning-Detection* (TLD) [5] algorithm allows real-time tracking of objects. The object of interest is defined by a bounding box in one single frame and without any previous training.

Because it is advantageous for the prototyping when results can be reproduced, a ROS node that feeds an image sequence into the system is used in the input layer. A network of ROS nodes, which perform different sub-task within the attention system, populates the attention layer. The output layer contains nodes to convert the object-based FOA into the bounding box and the tracker itself. The components used in this example are marked with “+” in fig. 1.

The purpose of this setup is the investigation of methods to use the attentional output to initialize the tracker. Research questions include: how to obtain a bounding box, when to start tracking, and when to stop tracking and consider a new FOA.

### 4 Outlook

Future work will be directed at integrating top-down influences. Many attention systems offer rudimentary support for these. However, feedback connections (gray arrows in fig. 1) from the task to earlier layers (including possible adjustment of preprocessings in the input layer) have not been sufficiently investigated.

Moreover, the flexible environment suggested here allows the definition of a framework for strict quantitative testing: On the one hand, the attention layer may be encapsulated by a system that provides input and compares the results to ground truth; in this way different attention algorithms can be compared. On the other hand, such an evaluation environment could also encompass an attention and a task layer, allowing to operationalize the performance of attention systems via a high-level tasks.

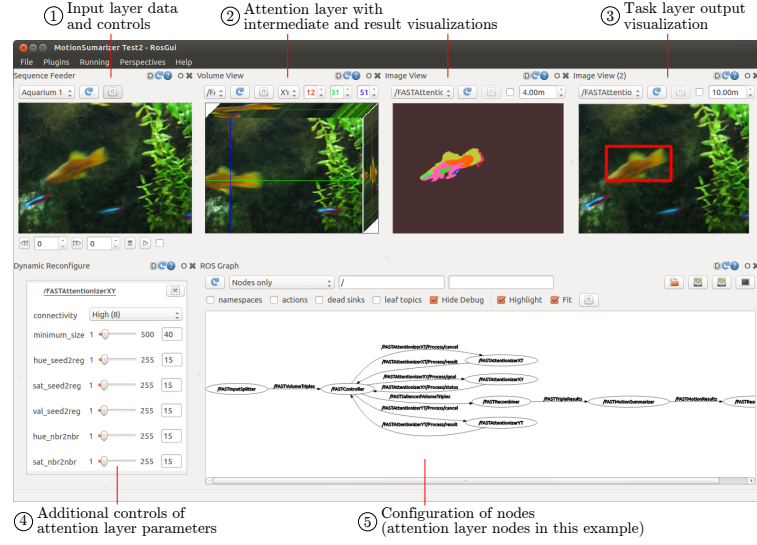


Fig. 2: ROS user interface showing a test configuration of the proposed environment. An image sequence is fed into the system ①, processed by the model reported in [10] ②, and subsequently forwarded to the object tracker proposed in [5] ③. Additionally, ROS tools for parameterization ④ and system visualization ⑤ are included.

## References

1. Aziz, M.Z., Mertsching, B.: Fast and robust generation of feature maps for region-based visual attention. In: IEEE Trans. Img. Proc. vol. 17, pp. 633–644 (2008)
2. Gazebo Wiki: <http://gazebo.org/> [online; accessed 10-july-2013]
3. Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In: IEEE CVPR. pp. 1–8 (2007)
4. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. In: IEEE PAMI. vol. 20, pp. 1254–1259 (1998)
5. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. Pattern Analysis and Machine Intelligence 34(7), 1409–1422 (2012)
6. Kotthäuser, T., Mertsching, B.: Validating vision and robotic algorithms for dynamic real world environments. In: Simulation, Modeling, and Programming for Autonomous Robots, pp. 97–108. Springer (2010)
7. Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., Ng, A.Y.: ROS: an open-source robot operating system. In: ICRA workshop on open source software. vol. 3 (2009)
8. ROS repository: saliency\_detection, [http://ros.org/wiki/saliency\\_detection](http://ros.org/wiki/saliency_detection) [online; accessed 10-july-2013]
9. Rudinac, M., Jonker, P.P.: Saliency detection and object localization in indoor environments. In: IEEE ICPR. pp. 404–407 (2010)
10. Tünnermann, J., Enns, D., Mertsching, B.: Saliency-guided perceptual grouping using motion cues in region-based artificial visual attention. In: IJCAI Workshop ISACS (2013)
11. Tünnermann, J., Mertsching, B.: Region-based artificial visual attention in space and time. Cognitive Computation (2013)