

# A Modular Vision System for Cognitive Robotics Research

# icVision

J. Leitner, S. Harding, M. Frank, A. Förster, J. Schmidhuber

Dalle Molle Institute for Artificial Intelligence (IDSIA)  
Scuola universitaria professionale della Svizzera italiana (SUPSI)  
Faculty of Informatics, Università della Svizzera italiana (USI)



juxi@idsia.ch | <http://www.idsia.ch>

## Abstract

We present an easy-to-use, modular framework for performing computer vision related tasks in support of cognitive robotics research on the iCub humanoid robot.

The system performs object recognition and localization of objects in 3D space.

One of the main design goals is to allow rapid prototyping and testing of vision software on the iCub and reduce development time by removing redundant and repetitive processes. The framework is implemented in C++, and uses YARP and HTTP for connectivity and OpenCV for the underlying image processing.

## The icVision Framework

The framework consists of distributed YARP modules interacting with the iCub hardware and each other. At the center is the **icVision Core** module, which handles the connection with the hardware and provides housekeeping functionality (e.g., extra information about the modules started and methods to stop them).

Other modules can attach to the **icVision Core**. Currently implemented modules include object detection, 3D localization and saliency maps.

These are reachable via defined interfaces, which allows easy swapping and reuse of other YARP modules (e.g. iKinHead and stereoCam both provide pixel to 3D location conversions, and can be used instead of the provided one). Some modules also expose their functionality by HTTP, allowing for rapid prototyping without the need for YARP.

## icVision Filter

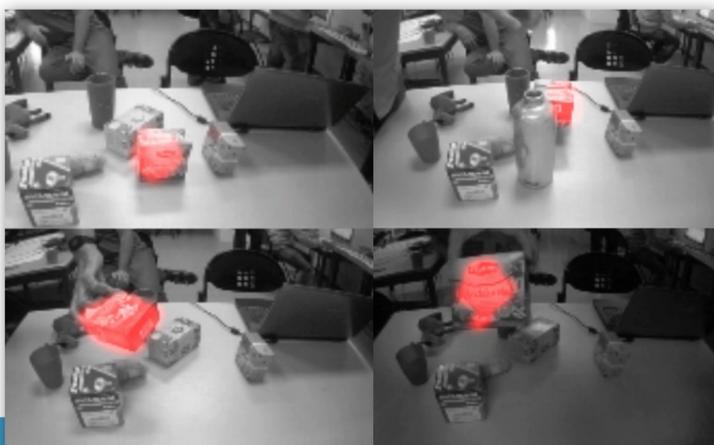
The **icVision Filter** module template provides easy access to the iCub's camera images (removing the overhead of reimplementing the YARP port communication) and simplified bindings to the OpenCV library. Furthermore this module provides testing functionality to allow the vision system to run without having YARP installed or being connected to a live robot. Using files as inputs, the user can develop object recognition and tracking algorithms offline.

The example below shows a very simple image filter that shows only the bright red objects within the iCub's field of view. Using machine learning, more complicated filters can be generated automatically. The sequence of images below shows an object being tracked in real time using a learned filter.

```
// Author: Juxi Leitner <juxi.leitner@gmail.com>
#include <icVision/icFilterModule.h>

class RedDetector : public icFilterModule {
protected:
    void setUsedInputs() { usedInputs.Add(1); }
public:
    RedDetector() { setName("RedDetector"); }

    icImage* runFilter() {
        icImage* in = InputImages[1]; // red
        icImage* out = in->threshold(64); // cvThreshold
        return out;
    }
};
```



A icVision Filter module running and tracking a tea box  
Video: <http://youtu.be/37YwAuY35gk?t=2m1s>



Placing an object in the world model



Video: <http://Juxi.net/projects/icVision/cs.mov>

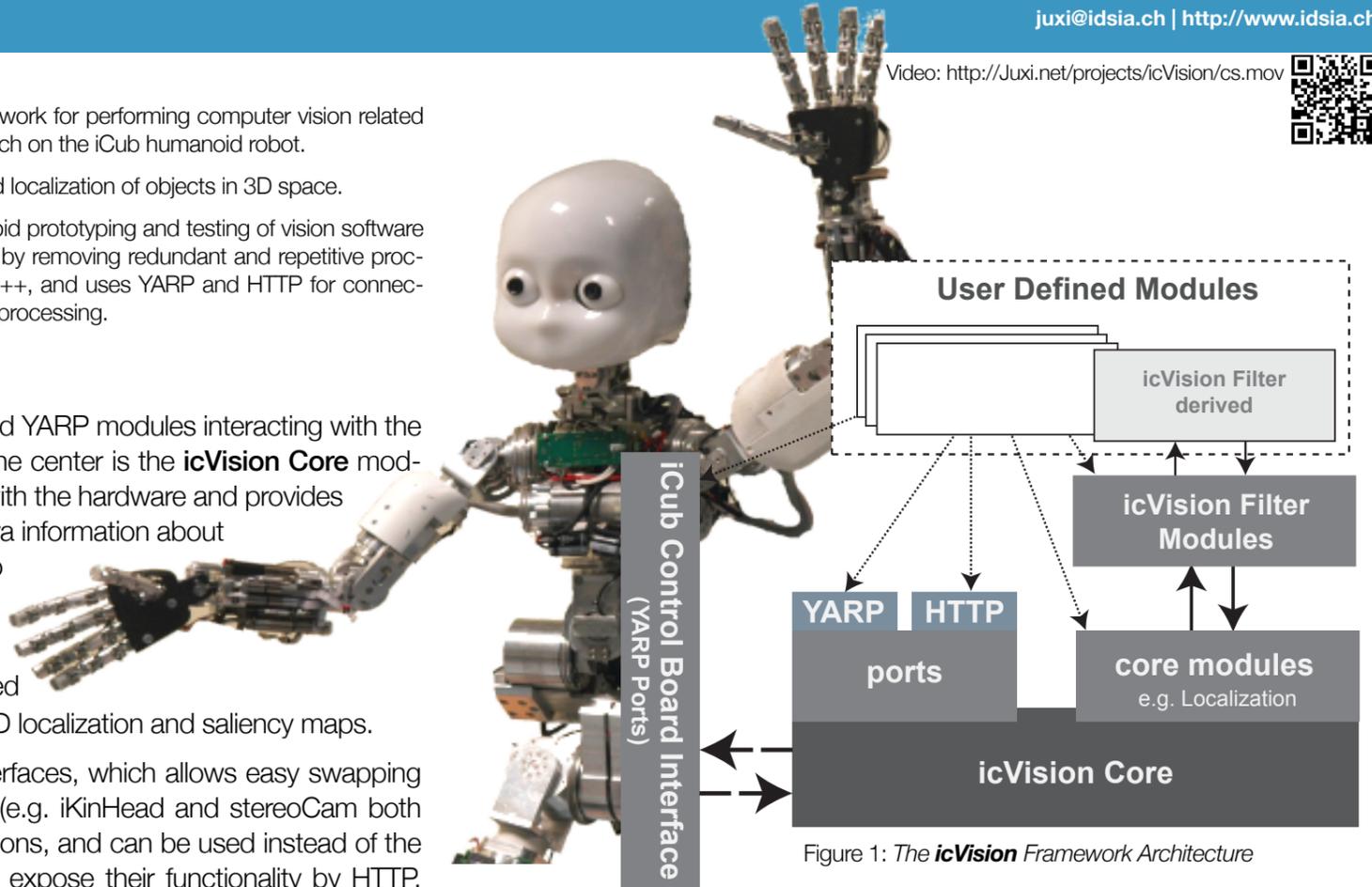


Figure 1: The **icVision** Framework Architecture

## icVision Localization

The **icVision Localization** module is one of the modules provided by the core framework. It allows for conversion between camera image coordinates and 3D coordinates in the robot reference frame. Using the objects location in the cameras (provided by an **icVision Filter** module) and pose information from YARP, this module calculates where the object is in the world. This information is then broadcasted via YARP for other modules to use, e.g. a roadmap planner or a grasping subsystem. Below the localized object can be seen after being placed in the world model.

## Current Research Involving icVision

The IM-CLeVeR project [1] aims to develop new robot controllers based on the principles of autonomous development. Cognition and perception are seen as the foundation to developmental mechanisms, such as sensorimotor coordination, intrinsic motivation and hierarchical learning. We are trying to extend the machine learning and computer vision research onto the iCub platform.

To perceive, **detect and track objects**, and help building a world model, which is used for tasks like motion planning and grasping.

Realtime, **incremental learning** is applied to further improve perception and the model of the environment and the robot itself.

Learning to grasp and **basic hand-eye coordination** are other areas of research this framework is currently applied.

## Next Steps

Research towards Human Robot Interaction: e.g. gesture recognition [3], handwriting recognition [4] on the iCub.

Creating new object filters by stacking various filters, allowing for hierarchical and incremental computer vision.

## References

- [1] The IM-CLeVeR project: Intrinsically motivated cumulative learning versatile robots, *Baldassarre et al.* (2009)
- [2] A Research Roadmap of Cognitive Vision, *European Research Network for Cognitive Computer Vision* (2005)
- [3] Max-Pooling Convolutional Neural Networks for Vision-based Hand Gesture Recognition, *Nagi et al.*, Conf. on Signal & Image Processing and Applications, (2011)
- [4] On Fast Deep Nets for AGI Vision. *Schmidhuber et al.*, Conf. on Artificial General Intelligence (2011)