STEREO VISION WITH COGNIMEM

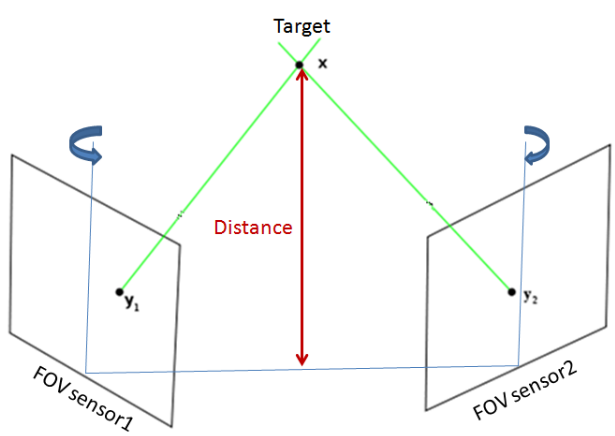
THE APPLICATION

Stereo Vision is critical for the evaluation of the proximity of an object and is the starting point for many machine vision applications. Several commercial systems are available and produce accurate results, but they are bulky and costly which prevents their use as a commodity in eye-operated communication devices, smart motion sensors, robotic guidance, etc. The CogniMem pattern recognition chip can be an enabler for a miniature and low-cost stereo tracker due to its recognition speed, low-power consumption and abilities to match fuzzy and ill-defined objects and perform real-time learning .

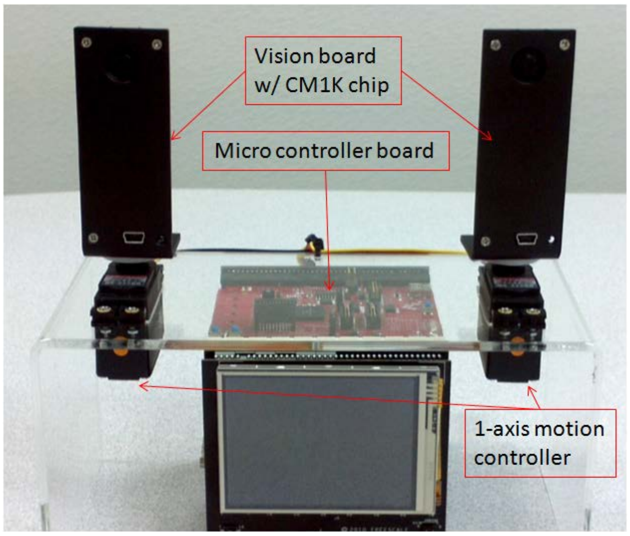
This application note describes how to implement a stereo tracking system based on the CogniMem technology.

THE METHOD

A Stereo Tracking system is composed at the minimum of two sensors placed at a known distance from one another, a target recognition and tracking algorithm applied synchronously to the video images of each sensor and a triangulation formula to estimate the distance of the target to the plane common to the sensors' axes.



Example of a setup with 1-axis motion



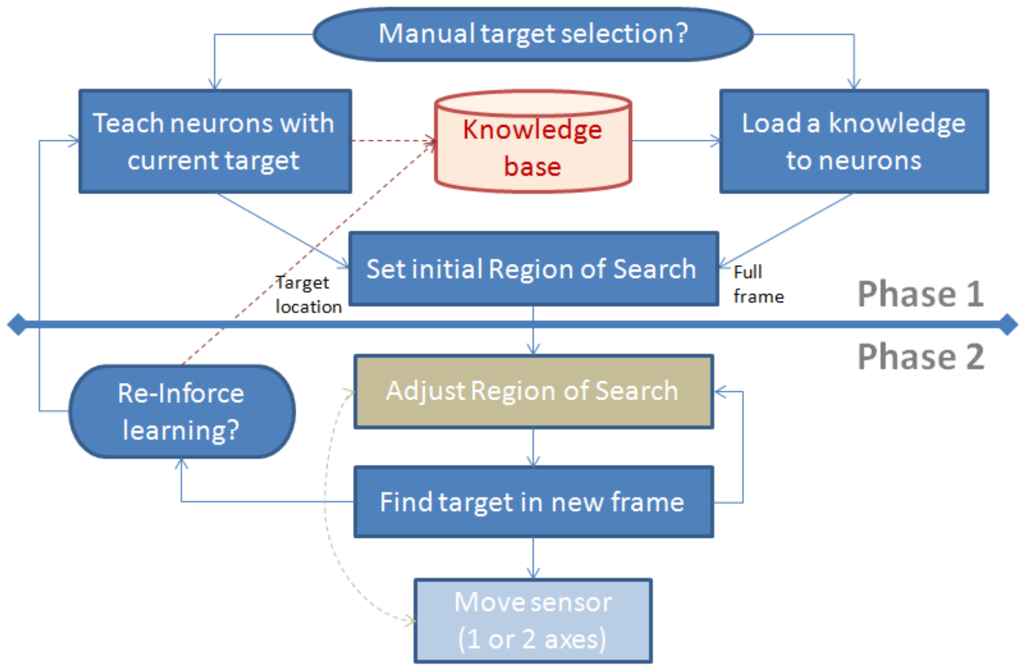
The stereo vision algorithm can be decomposed into three phases as described below and the CogniMem (CM1K) chip offers essential benefits in each phase.

PHASE 1: IDENTIFICATION OF THE INITIAL TARGET (REAL-TIME OR DEFERRED)

- Depending on the context of the application, the initial target identification can be made manually by an operator or automatically by a target identification engine.
- In the first case, assuming that the operator can select the target on a screen, the neurons of the CM1K are immediately put to work and automatically learn the features of the target. The process of knowledge building has started and recognition is enabled. Note that the manual learning can be executed on one of the two cameras only. As soon as the initial target is detected then it must transmit its location to the second camera.
- In the second case, the features of the target to detect can be retrieved from a knowledge base built ahead of time and saved to a file or stored on board a Flash memory or else. The neurons of the CM1K can load this knowledge in their memory and start recognition immediately. If the final system involves two identical vision boards, the CM1K of both boards have to be loaded with the same initial knowledge.
- The feature vectors to extract from the location of the target can be multiple and contain information describing to the shape, texture and/or color of the target learned as a whole or in parts.

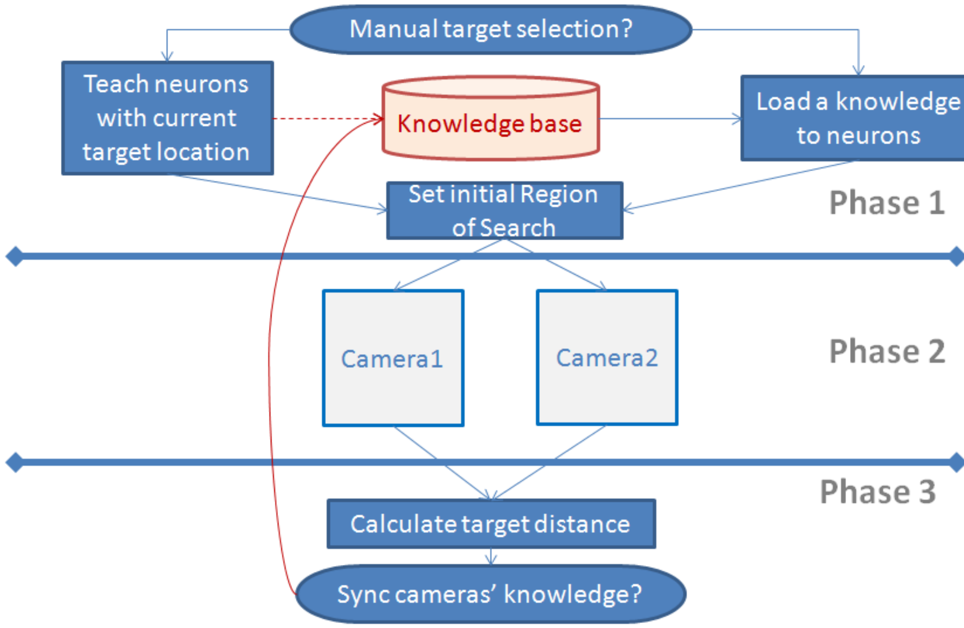
PHASE2: REAL-TIME TRACKING AND ADAPTIVE LEARNING

- Use of the high-speed recognition capabilities of the neurons to find the target and calculate the smallest possible region of search around it in the next frame.
  - If the vision sensor is mounted on a motion controller, the region of search is changed mechanically by moving the sensor so the target is re-centered in the field of view.
  - If the vision sensor is mounted on a fixture, the region of search is changed programmatically.
- In the event that the confidence of the neurons decreases because the target is changing and its initial features become less and less relevant, the neurons will automatically learn the latest features of the target to keep tracking correctly. This is called real-time reinforced learning and will occur if the target gets closer or farther, rotates, gets partially hidden or in the shade, etc.
- Since the two cameras do not see the target in the same focal plane, they can start expanding their common initial knowledge base with different feature vectors or the application controller can ensure that the latest features are learned by both cameras.



PHASE 3: DISTANCE EVALUATION

- For accurate calculation of the distance of the target to the plane defined by the two cameras, both vision sensors must be triggered to grab a new frame at the same time.
- Provided that the target has been recognized by both cameras, its location in the two fields of view is submitted to a simple triangulation formula returning the distance of the target to the cameras.
- The resulting distance can be used to close several loops simplifying the recognition and consequently the tracking mechanism:
  - Load of a different “primitive” knowledge base based on the scale of the target in the field of view and optimization of the region of search for the new scale.
  - Adjustment of an auto-zoom lens



THE CHALLENGE

The challenging aspects of a stereo vision system depend on the type of target to identify, the scenery in which it will evolve, its speed and behavior. They can be summarized with the following two action items:

- Build an initial primitive knowledge capable of detecting the target when it enters the field of view of a camera knowing that its scale, orientation and color intensity are unknown. Note that this primitive knowledge can be a set of multiple knowledge built on a variety of features and contexts.
- Define a tracking algorithm minimizing the region of search around the target without risking losing it. Indeed, the larger the region of search, the slower the response time and higher the probability that the target exists the region at the next frame.

THE PROPOSED SOLUTION

- A miniature system with the following components:
  - CMOS sensor with global shutter, external trigger
  - Lens (optional auto-zoom)
  - CM1K chip with 1024 neurons (enough to hold a primitive knowledge and expand it during re-inforced learning)

- Processor or FPGA to synchronize the operations between the sensor, the neurons and the motion controllers (whether programmable regions of search or external motors). Also establish communication with the other camera and outside world.
- Communication port (choice of Wi-Fi, USB, Ethernet, etc)
- Miniature motion controllers