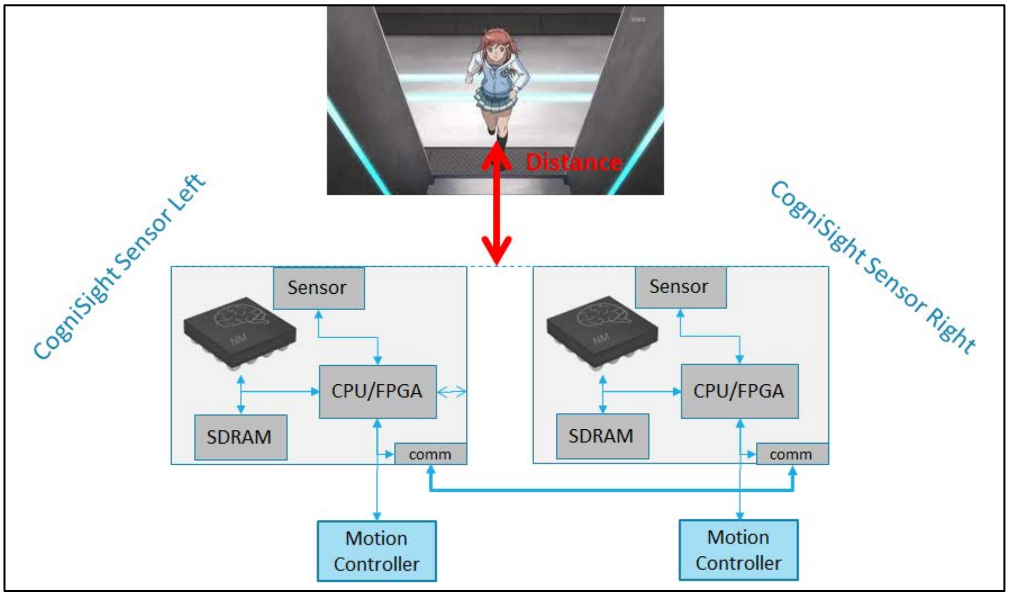
Stereoscopic Target Tracking

Stereo Vision is critical for the evaluation of the distance 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 NeuroMem chip can be an enabler for a miniature and low-cost stereo tracker due to its recognition speed, low-power consumption, ability 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 a CogniSight sensor which can be described as a device combining a video input, a CPU or FPGA and a NeuroMem network as illustrated below. The CPU or FPGA oversee the video acquisition, as well as the image learning and recognition logic specific for active stereoscopy and described in the application note.



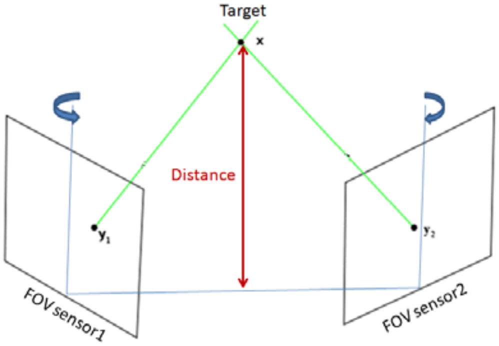
View videos on GV youtube account

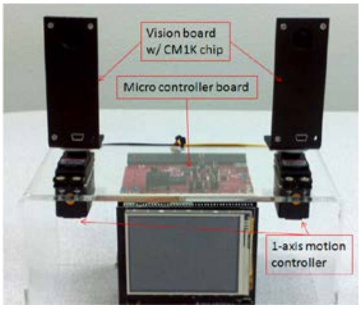
- Stereo Tracking with two motorized cameras
- Air Plane tracking with motorized camera
- Car tracking with fixed camera
- Face tracking with fixed camera

Overview

A stereo tracking system is composed of two CogniSight sensors placed at a known distance from one another and sharing a same vertical plane. The initial target can be learned in either Field Of View (FOV P) and after each learning operation, the knowledge of the sensor P must be copied to the other sensor. This ensure that both sensors use the same knowledge to recognize the target.

After each frame, each sensor has a map of the locations where the target is recognized and can produce a center of gravity, integrating or not the confidence level of the firing neurons. One of the sensor can oversee the calculation of the final triangulation formula to estimate the distance of the target to the plane common to the sensors' axes.





Setup featuring 2 V1KU cameras with 1 Aptina CMOS, 1 Actel FPGA and 1 NeuroMem CM1K chip per V1KU

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, if the operator can select the target on a screen, the neurons 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 can load this knowledge in their memory and start recognition immediately. If the final system involves two identical vision boards, The NeuroMem network of both boards must 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 in 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.

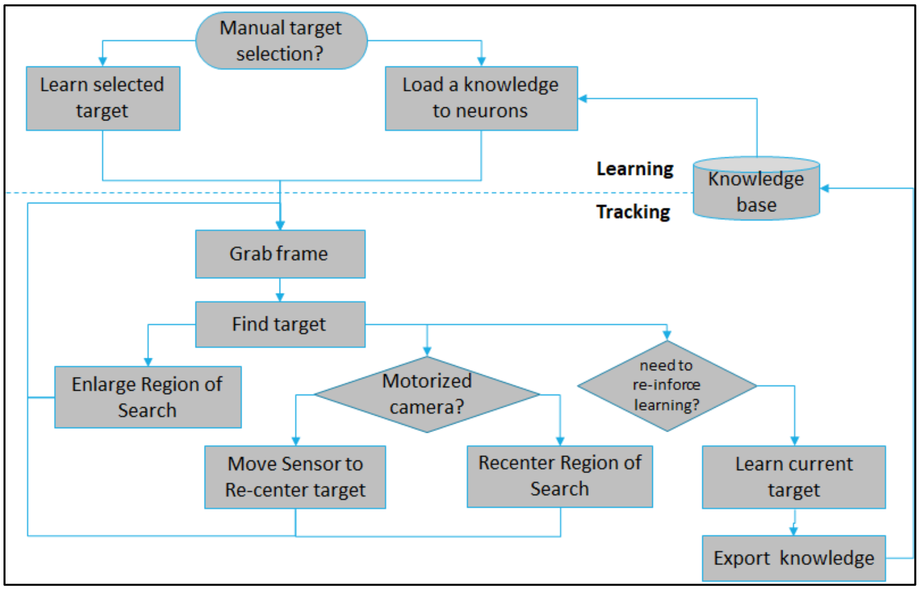
- If the vision sensor is mounted on a motion controller, the region of search is changed mechanically by moving the sensor so the target is re-centered in the field of view.
- If the vision sensor is mounted on a fixture, the region of search is changed programmatically.

The challenge in tracking is to define how to minimize 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.

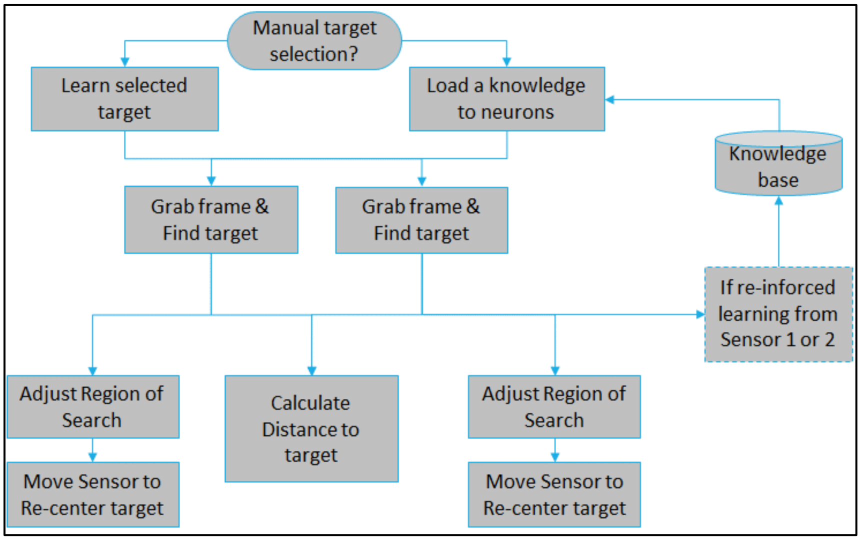
If the confidence of the neurons decreases because the target is changing and its initial features become less and less relevant, the neurons will automatically learn the latest features of the target to keep tracking correctly. This is called real-time reinforced learning and will occur if the target 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 both cameras learn the latest features.

Same tracking algorithm per sensor



Synchronization between to sensors



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 next recognition or allowing hypothesis generation including:

- Load of a different “primitive” knowledge base based on the distance to the target to lock the tracking onto finer details for example, or on the contrary re-lean the overall shape of the target which is scaling down.
- Adjustment of an auto-zoom lens
- Estimating the next position of the target knowing its speed
- Etc.