Sightech Vision Systems, Inc.

Trainable Vision

How Intelligent Vision Systems View the World

Self-learning vision, animal-based or not, is impressively useful because it automatically assimilates massive amounts of information from the world and then, with this information, can perform impressive navigation, detection and inspection tasks.

Published 2005

Sightech Vision Systems, Inc.
6580 Via del Oro
San Jose, CA 95126
Tel: 408.282.3770 Fax: 408.413-2600
Email: sales@Sightech.com
Web: www.Sightech.com

Table of Contents:

The Importance of the way intelligent systems See the World

Different animals see the world around them in different ways. Some see only grayscale and not colors. Other animals can see very small objects – such as raptors do from high in the sky. Other animals are very keen on detecting the slightest movement. Animals have evolved so they apply their visual skills in specialized ways that benefit them the most. They usually do not waste their mental bandwidth on visual capabilities that they do not critically need to be successful.

The same is true of self-learning technology. Since we can choose the way our system sees the world, we can improve its level of competence for any given task. We will focus, for animal or machine, the "view" of the world.

In processing vision, animals reduce data to "view" their world using at least steps

a) Receive the image at the retina – the initial data is at the "pixel" level.
b) Reduce by simplifying a pixel of image data to an RGB (single color) representation at the retina receptors. This is a large reduction because the actual data is "hyper-spectral" (complex spectral composition), and the resulting data is only 24 bits which represents one visible color.
c) Locally preprocess and reduce the image with interacting localized neural processes at the retina level.
d) Further reduce and transport the image data to other neural functions via the optic nerve.
e) Move reduced data towards the brain. Mental processes there further reduce the data.
f) Drive and use a neural conceptualization process, where concepts (objects) are learned and/or triggered.
g) Make an aggregate decision for action.

Trainable vision performs many of the same functions

a) Data is presented at the lowest level by the imaging device (camera).
b) Image pixel is simplified to a single RGB color – 24 bits.
c) This data is then immediately reduced by a multitude of what we call "Views". These "Views" are guided by what we call a "Feature". A "Feature" defines what the world "looks like" when observed by the views. In other words, the Feature defines how the trainable vision system actually organizes and "sees" the world at a lowest-level.
d) The data from all of these Views (guided by the Feature definition) is fed further into the learning system for learning (conceptualization and generalization) and recognition purposes.
e) After training is accomplished, the trainable vision system may be operated. When operating the vision, the data processed in the same way, and actual results are produced . The vision system can detect all kinds of defects and absence/presence of objects.

f) Operation employs a sophisticated decision process that consists of a four level hierarchal voting process that arrives at an aggregate decision of PASS/FAIL. The user has parametric control over each and all of these decision hierarchies.

Kinds of input image data that can comprise a "view"

a) Gray-level – this is an 8-bit value per pixel that represents the intensity.
b) Binary image – this is a black-or-white (binary) representation of the image. Edge convolutions and comparing the image, pixel-by-pixel, to an intensity threshold are common way to create a binary image.
c) Color image data. PC-Eyebot uses 24bit RGB format.
d) Behavior – image movement from frame-to-frame (coming soon).
e) Multi-angle – multiple camera images from different perspectives (future development). 3D vision is in this category.
f) Hyper-spectral – this is a complex spectrum of wavelengths are collected for each pixel – not the highly simplified RGB pixel of data.
g) X, Y, XY, R, etc. This location data may be mixed with other data to comprise a very effective View that "cares" about locations as well as associations.

In summary, in order to see the "big picture", animals and machines both MUST see the "small picture"

Animals and trainable vision obtain high-level decision from a large visual field-of-view. This involves usually from 300k to several million pixels of data. The number of combinations possible in an entire image is beyond galactic in size – this image will never be visited again in the live of the universe. This fact requires the real world and its large number of pixels be observed by animal (or machine) first at a low level. A multitude of low-level input then must undergo severe data reduction and aggregation. Learning processes combined with hierarchies conceptualize and reduce the data further. The result of this large massive process is the high-level detection of an object, or in the case of inspection, an aggregate decision. This usually results in some sort of favorable action.