A white paper from General Vision

# Trainable vision sensors, better than smart cameras

The trend of the past years has been to childishly label cameras as smart because of a high-speed network connection or a powerful processor running complex algorithms. Still, these cameras remain pretty hopeless in front of the amount of data they generate and do not respond to real needs for flexible and accurate image recognition with realistic speed performance.

Indeed the market has been ready for cameras which are smarter, smaller, faster and cheaper. This can be translated as follow:

- The sensor must have the ability to learn and recognize patterns autonomously
- The sensor can be taught directly by the user by showing examples and without any programming.
- The sensor can transmit results and/or images only when information of interest is detected.
- The sensor must be small for portability and cost reduction.

The trainable vision sensors developed by General Vision and branded as the CogniSight sensors respond to the above criteria and make image recognition ubiquitous because easy and affordable.

## CogniSight sensors merge vision and decision in a miniature apparatus

A CogniSight sensor features a video sensor and a brain which can learn, make instantaneous decision and selective transmission. This brain is trained and controlled by CogniSight, an image recognition technology developed by General Vision and based on multiple neural networks.

The first time you install a CogniSight sensor, you will have to go through the few necessary steps common to all camera installation: mounting at the correct distance from the objects, lighting for a good contrast, connection to a photocell or triggering device if applicable and finally interface to a PC for minimum configuration. This is where the roads between a CogniSight sensor and a smart camera split. The CogniSight sensor is equivalent to an image recognition server. It operates autonomously and just needs to receive the initial instruction from a client application to know what to recognize and which actions to take accordingly. This instruction can be reduced to the transmission of a knowledge file to the sensor which is equivalent to loading a knowledge into its associative memories. The knowledge is a simple data structure defined within the CogniSight technology. It is built by learning examples, can be extended and fine tuned at any time, saved to file for portability and read by any sensor powered by a CogniSight recognition engine. The client application connected to the CogniSight sensor can be designed to let a user perform immediate interactive training and/or to simply load existing knowledge files.

## CogniSight sensors learn by examples and are quick at picking up patterns

Without programming any line of code, nor reading image processing manuals, you can teach a CogniSight sensor what to do just like you would teach a new operator. Simply outline your objects of interest in live or still images and label them with a category. On a factory floor, you can classify parts as Good or Bad; Acceptable, Defective or Recyclable; Type1 to Type N. Inside a video surveillance station, you can classify and count passing objects as “car”, “truck”, “van”, “bicycle”, etc. As soon as you teach one example, the sensor commits a neuron and starts classifying incoming objects with this start-up knowledge. At first, it will probably classify all the objects as belonging to the one and only known category. If the features seen in an image are really different from the one learned by the sensor, the response can be “Unknown”. This will continue until you teach something new. As soon as you teach a second and other categories, additional neurons get committed and they all start delimiting their similarity domain. The more you teach, the more neurons and the less confusion between different categories of objects.

## Immediate recognition and decision

Since the CogniSight sensor can recognize what is present in an image, might as well let it take the appropriate action based on what is recognized in the image. On a factory floor, the sensor can directly control an ejector to reject or divert parts, transmit periodic statistical report about the production and, if needed, record the images of parts with a certain type of defect. On a surveillance station, the sensor can send a warning signal, record the date and time a special event occurs and transmit each image showing a particular object.

Once a CogniSight sensor is trained, it can save its knowledge for later use. This means that the second time you will power the sensor, it can automatically and autonomously resume inspection with no need for a PC or network connection. You can always connect a PC if you wish to monitor production statistics, start a new training or add more training to an existing engine. These interruptions are made over TCP/IP and do not interfere with the real-time recognition running between 15 to 20 frames per second.

## So simple, yet so flexible and adaptive

The learning mechanism of a CogniSight sensor is very intuitive since you only have to show and label examples. This “trainability” makes it possible to transfer to the sensor a human expertise built on years of practice, with no need for intermediaries such as interpreters, consultants and programmers. Still, there is no unique way to train a sensor to recognize a family of objects and you may have to teach and evaluate several engines before retaining the one which will be used for the real application. Let’s take the example of defect detection. Because it is impossible to model all the possible representations of defective parts, you can try two teaching methods: (a) be very conservative at teaching good parts and consider everything slightly different as defective. (b) be moderate at teaching defects so that more or less similar defects can be picked up. The selection of a teaching method can be influenced by the frequency of occurrence of the events or objects to recognize. If your production has 0.02% of defects, it will be difficult to capture these defects and a conservative teaching might be wiser..

## A validation driven by the cost of a mistake

Prior to letting the sensor take actions based on the contents of images, a thorough validation of the recognition engine is critical. This is simply done by monitoring the response of the engine over a significant number of examples. In a production environment, this verification may span over several days to ensure that the sensor has seen and classified properly a representative number and diversity of products. If corrections are required, they can be taught immediately by the end-user. Unlike in conventional imaging systems, correcting the engine does not require to fix bugs in a program, but rather to verify that the taught examples were correct and detailed. Indeed the CogniSight technology has been used and field-tested for several years and its classifier is endorsed in many thesis and applications. The key to the accuracy of a recognition engine is to have good teachers.

Depending on the cost of the error for an application, it might be necessary to publish strict guidelines for the teachers and even require that multiple experts annotate the same objects to make sure that there is no possible error of classification. If two human experts cannot agree to the classification of an object, you cannot expect wonders from a CogniSight sensor which acts as a simple parrot.

To help test a recognition engine, the CogniSight sensor can transmit images at high speed and selectively if requested. The CogniSight client application can then save these images and even dispatch them into folders per category. On a manufacturing line, you can collect images of products during different time and crew shifts. In an outdoor surveillance station, you can collect examples of objects under a variety of lighting and weather conditions, etc. The ability to request that the sensor transmits only certain categories of images can greatly speed up the validation of a recognition engine. For example, if you can quickly notice if the images stored in the folder "Category A" really belong to this category or not. If so, you can assume that the engine is well trained to identify category A. On the other hand, you will have to verify that images of category A cannot be found in the other folders. Now, if you notice that images stored in "Category A" often belong to category C, you can immediately teach more examples of objects belonging to category C. These observations help you identify the categories of objects which require more teaching due to the complexity or diversity of their features. You can also easily locate the categories which introduce confusion and need to be taught with a conservative approach. These observations can also lead you to define intermediary categories or, on the contrary, to merge categories together.

## Building image knowledge bases

On many occasions, it is more practical to teach and validate a recognition engine off line, that is without the sensor. You can take all the necessary time to teach and screen the recognition results, understand where confusion occurs, try new teaching methods and qualify recognition performance in term of throughput and accuracy. Also, sometimes, remote training might be necessary to consult an expert, or to compare the teaching instructions of multiple operators. Again, training is the key to the accuracy of an engine. The images collected with the CogniSight sensor can be used to build several knowledge files. The Image Knowledge Builder application from General Vision is intended for this purpose. It can build a knowledge from a series of annotations made by an operator and test it on many images. Diagnostics reports guide you in the process of compromising between throughput, accuracy and speed. For example, an engine cannot discard what it has already learned. Image Knowledge Builder can pinpoint however when a knowledge starts to degenerate due to inconsistencies in the set of training examples. When satisfied with a recognition engine, you can simply save it to an Image Knowledge (IK) file and load it on the sensor.

## Adapting a knowledge to changes and variations

A CogniSight sensor can learn many models of objects and solve recognition problems with high non-linearity. For example, if it is easy to model a good screw from a bad screw, it is more tedious to describe a good fish from a bad fish, especially when the fish is not one among a dozen swimming in an aquarium, but rather one herring caught with a million of other herrings. A common challenge in industrial inspection is adaptivity to change of production, whether it is the introduction of a new part, the change of one component in a part, or else. With a CogniSight sensor, this problem can be solved in a very practical way: add more training to an existing recognition engine, or start a new recognition engine. For example, as ackward as it may sound , a CogniSight sensor trained on a fishing boat in December to sort herrings has a good chance to perform poorly in July. The explanation is very pragmatic: Fishes happen to loose their scales at different seasons and this directly alters their reflectivity. As a result, the images seen by the camera in July are different from the ones seen and taught in December. The customer had the option to build a "universal" engine recognizing the scale of herrings throughout the year, or to build several smaller engines for each season. Both options request a new image collection, some training and validation, and all these operations can be performed by the end-user.

Depending on the features of your objects, you might very well be able to teach a “universal” engine capable of classifying your entire production. This is especially true if you simply need to sort your parts into Pass and Fail. You can truly benefit from the non-linear classifier of the CogniSight sensor which will have no problem to store in a same knowledge examples of good low-fat strawberry drinkable yogurt, pine-apple-flavored cottage cheese quart and vanilla yogurt. Your only risk trying to merge all these models in a same knowledge is to end up with a large knowledge base (which is not necessarily a problem) and, possibly, a lesser accuracy. This can tested quickly with tools like the Image Knowledge Builder so you can make decision based on comparative results.

Another reason to start a new recognition engine can be un-related to the appearance of the objects to inspect but rather related to a level of moderation of the engine. For example, during hot summers, the quality of a label on a beer bottle might be less critical than during winter time. As a consequence a brewery might use two different engines, one being more tolerant to certain types of defects than the other, therefore increasing the throughput of the production.

## Smaller size and price tag

The evolution of the CogniSight sensors is going towards smaller, cheaper and faster appliances.

Smaller and cheaper makes it possible to multiply inspection stations along a production line. Instead of building one expensive system checking all the acceptance criteria of a part, multiple CogniSight sensors can be distributed along a factory line and perform simple inspection tasks with standard equipment (lens, lighting, ejectors, etc). For example, in the bottling industry, it makes sense to eject a bottle as soon as it is known that its filling level is below acceptance level, or it is not properly sealed or its label is improperly affixed. Taken individually, each of these controls are quite simple and prevent incoherence such as wasting supplies of caps and labels on empty bottles. Another advantage to early and local problem detection is the avoidance of damage and tear of the machinery.

The first CogniSight sensor was based on the ZiCAM CMOS unit from Pulnix and introduced in 1998. Its original dimensions were approximately 8”x4”x4” and its engine was composed of to 312 hardware neurons. The second generation of ZiCAM was a smaller unit (4”x4”x4”), but with still an expensive price tag of $6500. The latest CogniSight sensor is based on the Iris unit from Matrox. Its dimensions are 8”x4”x2” and its features high-speed Ethernet, 2000 software neurons. The next generation of CogniSight sensor will be based on the CogniSight-EB board developed by General Vision and with dimensions as small as 5”x3”, software and/or hardware neurons and a price under $2000 in volume. The main components of the CogniSight-EB are a video encoder, an ARM processor running Windows CE and ProAsic FPGA (Field Programmable Gate Array). Depending on the releases of the board, the CogniSight engine will reside on the processor or the FPGA. For OEMs, the board can be integrated into different enclosures to become either a single sensor system or a multiple-sensor junction box. The next evolution of the CogniSight engine will be a giant step for image recognition: a dual or single chip configuration leading to a size and cost ideal to make image recognition ubiquitous.

**For more information:**
General Vision
1150 Industrial Avenue, Ste A
Petaluma, CA 94951
www.general-vision.com
707-765-6150