Fundamentals of Machine Vision

BY PERRY WEST

AUTOMATED VISION SYSTEMS, INC.

AUTOMATED VISION SYSTEMS, INC.
San Jose, California 95124 U.S.A.
(408) 267-1746
www.autovis.com
2/1/2021

Table of Contents

Table of Figures

INTRODUCTION

Images are the inputs richest in information that man, animal, or machine can experience. There's little wonder, then, why people are pushing technology to make images work for automation.

After reading this white paper, you will have insights into machine vision, where it is used, how it benefits users, the different kinds of machine vision systems, and a very basic understanding of how vision systems work.

The discussion starts with the definition of machine vision and points out where the widening adoption of the technology may need modification to the definition. It illustrates the wide range of industries making use of machine vision, how they benefit from it, and the things that vision systems can do and how they can be used. It illustrates the different kinds of images used in machine vision, explores the anatomy of a vision system, the parts that every vision system needs to have. Finally, it looks into how the image is processed ending with a quick look into how deep learning is changing the way vision systems are programmed.

THE DEFINITION OF MACHINE VISION

For a long time, I used the following definition for machine vision:

The automatic acquisition and analysis of images to obtain desired data for use in controlling a process or activity.

Let’s look at the key words and phrases in this definition.

- Images – Very naturally, anything with “vision” must involve images. Images are central to vision whether human or machine. As we will see a bit later, there are a range of different kinds of images used in machine vision.
- Acquisition and Analysis – The machine vision is not concerned with acquisition alone, that would be photography. It is not concerned only with analysis, that would be computer vision. Machine vision involved both acquisition and analysis, often at high speeds.
- Automatic – This characteristic, that of being self-acting, is still dominant among machine vision systems. However, there are also vision system that are triggered by some event such as the press of a button or a voice command. It may be that this condition needs to be removed from the definition.
- Desired Data – Without life’s experiences, no technology, no matter how advanced, can evaluate an arbitrary scene and identify everything in the scene. To be feasible, the vision system must be expecting a “known” object and be able to return the specific needed data about that object.
- Controlling a Process or Activity – Early machine vision systems were used with machines, like robots, or on production lines to perform quality inspections. While these activities are still the majority of the activities for vision systems, there are uses outside these boundaries, such as counting people or providing information to aid in disease diagnosis, that suggests this limitation needs modification.

Finally, even the term “machine” in machine vision may need revision since there are many vision systems that do not work with or as part of a machine. Some people have suggested dropping the term “machine” and just saying “vision system”.

INDUSTRIES AND BENEFITS

INDUSTRIES

Machine vision touches a great many industries and activities. It is used by virtually every manufacturing industry of discrete parts. It has grown into other non-manufacturing areas, also. Table 1 below is a (partial) list:

Agriculture	Food/beverage	Packaging	Security
Auto manufacturing	Forestry	Paper	Self-driving cars
Biomedical	Gesture recognition	Pharmaceutical	Semiconductor
Container	Glass	Plastics	Solar
Cosmetics	Laboratory automation	Primary metal	Sports analytics
Display	Law enforcement	Printing	Telecommunications
Electronics	Logistics/warehousing/postal	Retail	Tobacco
Fabricated metal	Medical/healthcare	Rubber	Textiles
Fastener	Military/aerospace	Science	Wood products
Finance/Banking			

Table 1 -- Industries Using Machine Vision

BENEFITS OF MACHINE VISION

In the early days of machine vision, it was justified based solely on labor savings because that was the easiest to quantify. As understanding grew, other benefits were realized and enlightened management learned how to quantify them to help bring in machine vision to help their enterprises grow.

- Decreased scrap and rework – finding a problem early or, even better, being able to recognize when a problem will occur before it does is one way that machine vision aids in lowering costs.
- Improved productivity – the speed of vision systems allows production lines to run faster if humans do not need to keep up. There is less labor needed, and production line changeover occurs more quickly and smoother.
- Increased flexibility – the replacement of hard tooling with machine vision makes it practical to use a production line for a wider range of products.

- Better safety – machine vision can perform tasks that present safety risks to workers and allow the workers to be positioned where the risk to their safety is reduced.
- Better process control – along with whatever it is doing, machine vision can log useful production data that allows manufacturing management to spot trends early that might otherwise result in problems.
- Improve quality – machine vision inspection is more reliable than human inspection; it's free from boredom, fatigue, and distraction. Machine vision, being quantitative is more accurate than human judgement which is subjective and subject to variability. A machine vision system can be configured to make more inspections than would be practical for a human inspector.
- Reduced warranty and field service costs – when the quality of the shipped product increases, warranty costs and field service costs decrease.
- Improved customer relations – when higher quality products are delivered to customers, trust is enhanced and customers are more likely to place repeat orders.
- Enable new technology and cutting-edge products – many of the new products being introduced require abilities that humans cannot achieve.

APPLICATIONS AND FUNCTIONS

Although applications and functions of machine vision are complementary, we should differentiate them and show how they are related.

Functions are the capabilities of machine vision system; like make measurements (gauging). Applications are the purposes for which machine vision is used; like quality assurance. Below is a brief explanation of each application category and each function category. Table 2 and Table 3 below show how applications and functions relate.

APPLICATIONS

- Robot/machine guidance – a vision system provides the equivalent of hand-eye coordination as well as ensures a suitable environment. Examples are selecting the right part to pick up, precisely placing a part for assembly, and detecting obstacles or hazards that the robot or machine should avoid.
- Quality assurance – making sure the right part is assembled, that a part has the right dimensions, or there are no defects such as scratches on the part.
- Test and calibration – checking the visible output of a display to see that it is displaying the right information or looking at the input setting of a knob or dial so that the output can be verified.
- Real-time process control – checking the size, completeness, or correctness of a part or assembly at the point of manufacture so that adjustments or corrections to the process can be made immediately before rejects are produced.
- Data collection – providing data on the dimension, completion, or serial number for storage and later retrieval, perhaps to perform statistical process control.
- Machine monitoring – checking a machine, either the tool or the part, for dimensions or defects so that adjustments or maintenance is made immediately when needed.
- Material handling – unpacking, routing, or packing material.
- Sorting – separating discrete parts by attributes such as size, shape, or color.
- Counting – Keeping a count of object by attributes such as product code or when mass conveyed and a simple photo detector will not work.

FUNCTIONS

- Location and orientation – providing an object's pose, translation and rotation, in either two-dimensional or three-dimensional space.
- Gauging – provide dimensional measurements with precision.
- Flaw detection – detect an unwanted and unexpected element, like a scratch, that can only be described in generality and not with precision.
- Verification – check that an assembly has the desired parts and no undesired parts or a part has exactly the right number of features.
- Identification – reading a bar-code, 2D code, or human readable characters.
- Recognition – determination of a part's identity by its features such as size, shape, and number of holes.
- Tracking – following the motion of a part or guiding a tool, such as welding torch or glue dispenser, along a path.
- Safety – identify a condition that presents hazards to equipment or to workers and provide an appropriate alert.

Application	Functions
Robot/machine guidance	Location and orientation Identification Recognition Tracking Safety
Quality assurance	Gauging Flaw detection Verification Identification
Test and calibration	Location and orientation Gauging Recognition Tracking
Real-time process control	Gauging Verification Identification
Data collection	Gauging Verification Identification Recognition
Machine monitoring	Location and orientation Gauging Flaw detection Verification
Material handling	Location and orientation Gauging Verification Identification Recognition Tracking Safety
Sorting	Gauging Flaw detection Verification Identification Recognition
Counting	Identification Recognition

Table 2 -- Functions by Application

Function	Application
Location and orientation	Robot/machine guidance Test and calibration Machine monitoring Material handling
Gauging	Quality assurance Test and calibration Real-time processing control Data collection Machine monitoring Material handling Sorting
Flaw detection	Quality assurance Machine monitoring Sorting
Verification	Quality assurance Test and calibration Real-time process control Data collection Machine monitoring Sorting
Identification	Robot/machine guidance Quality assurance Real-time process control Data collection Material handling Sorting Counting
Recognition	Robot/machine guidance Test and calibration Data collection Material handling Sorting Counting
Tracking	Robot/machine guidance Test and calibration Material handling
Safety	Robot/machine guidance Material handling

Table 3 -- Applications by Function

IMAGES

Since images are central to machine vision, a good place to start is to look at what images are common in machine vision.

First, let's define what we mean by image, because there are two major types of images in machine vision: the optical image that is focused onto the image sensor by the lens and the digital image that is read off the image sensor and processed by a computer.

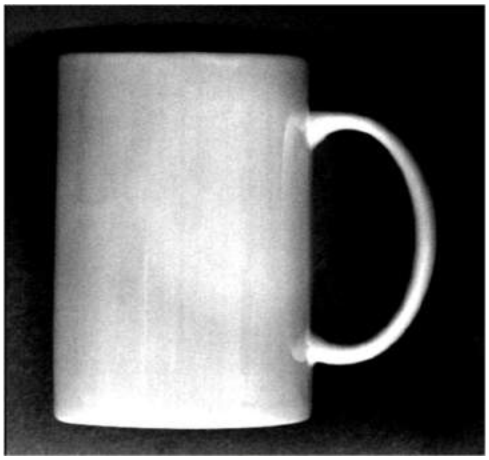


Figure 1 -- Optical Image of a Cup

OPTICAL IMAGE

The optical image (see Figure 1) is simply focused light energy – a representation or reproduction of light energy coming from the scene imaged. It is created on the image sensor. The digital image created by the image sensor cannot have more information, or detail, than what is present in the optical image.

DIGITAL IMAGE

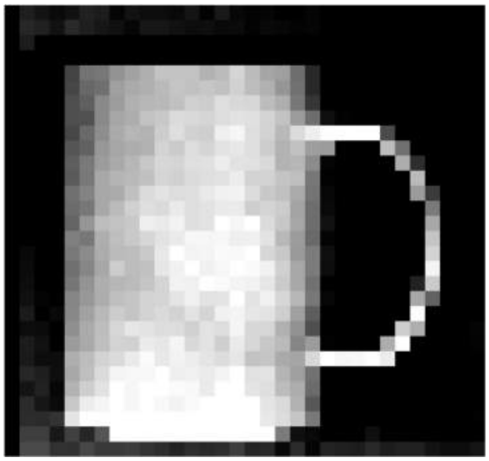


Figure 2 -- Digital Image of a Cup

The image sensor samples the optical image, both in space and intensity. In Figure 2 you can see that the image is sampled in both the horizontal and the vertical direction. Each sample is a pixel, which is short for picture element.

What is not obvious in Figure 2 is that each pixel has discrete values or gray levels. For most machine vision images, each pixel is represented by an eight-bit number and ranges in value from 0 to 255.

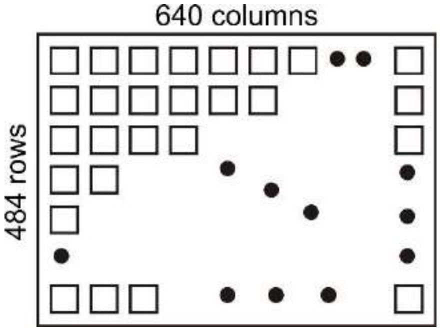


Figure 3 -- Image Resolution

The number of samples horizontally and vertically, or columns and rows of pixels is the image resolution (see Figure 3). A while ago, 640x480 pixels, or VGA resolution, about one-third million pixels or one-third megapixel, was the norm. With advances in image sensors, one to five megapixels is the most common range for image resolution, but image resolutions from VGA, one-third megapixel, up to 70 megapixels are commonly available.

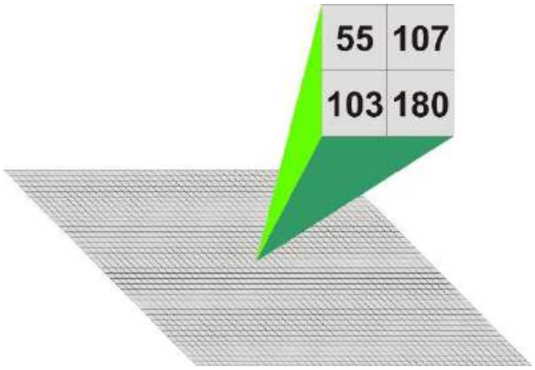


Figure 4 -- Gray-Scale Image

GRAY-SCALE IMAGE

The most common image in machine vision is the gray-scale image. In the gray-scale image, each pixel has a numeric value that is proportional to the light energy falling on the corresponding pixel in the image sensor as shown in Figure 4.

Most commonly, the value is eight bits giving values ranging from 0 to 255. Some vision systems use 10- or even 12-bit values giving ranges of 0 to 1,023 or 0 to 4,095. The image sensor and its circuitry have a noise component that limits the number of useful bits.

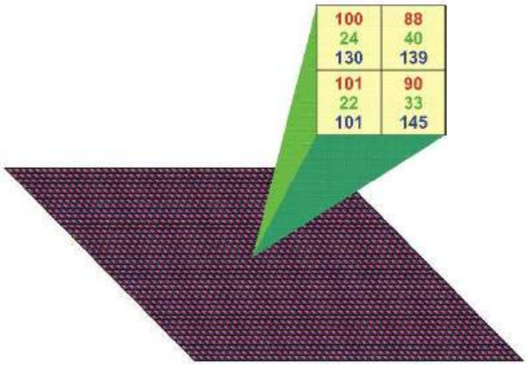


Figure 5 -- Color Image

COLOR IMAGE

The color image (see Figure 5) is different from the gray-scale image in that every pixel has three numeric values – one for red light, one for green light, and one for blue light energy.

BINARY IMAGE

A binary image, shown in Figure 6, is one a pixel can have only one of two values: light (usually represented by a “1”) and dark (usually represented by a “0”). Binary images do not come directly from a camera, and are created by image processing. A threshold value operates on the image leaving any pixel with a value below the threshold dark and all other pixels light. Thresholding, or binarization, is a method of simplifying the image that makes it easier to process. It does not work for all applications.

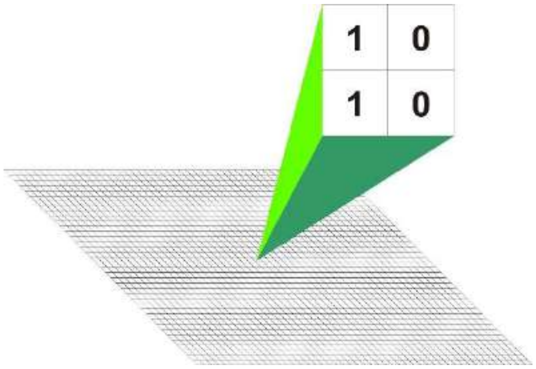


Figure 6 -- Binary Image

3D IMAGE

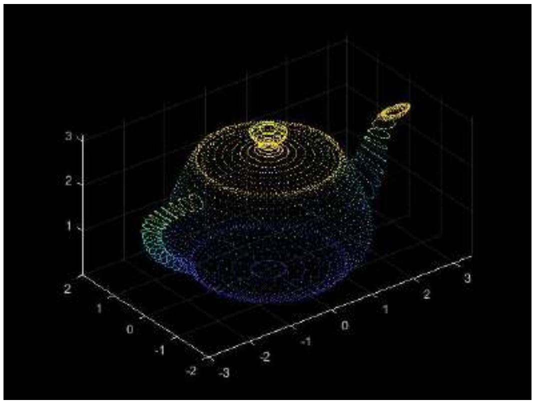


Figure 7 -- Point Cloud Image

While the use of 3D imaging systems is growing rapidly, this whitepaper does not cover this topic except in passing. There are over one dozen different technologies capable of creating a 3D image.

Most commonly, the 3D image is exactly like a gray-scale image. The difference is that the value of each pixel is a distance and not a measure of light level. There are 3D systems that supply both distance and light level values and have two values for each pixel. There are also 3D systems that capture a color image along with the 3D image. Often, the two images are kept separate, but if combined, or fused, the result is pixels with four values each.

Many 3D imaging systems do not provide a regularly spaced array of pixels, they supply a set of data points, each of which has an X, Y, and Z coordinate. This type of image is called a point cloud image as is shown in Figure 7.

ANATOMY OF A VISION SYSTEM

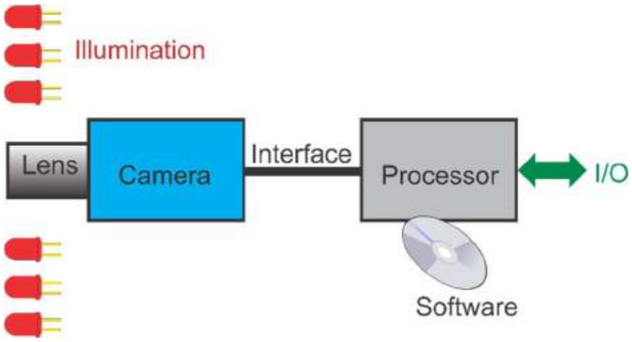


Figure 8 -- Diagram of a Vision System

It might be conjectured from the definition of machine vision that a camera and a processor combine to make a vision system. This is only partially complete. It is true a vision system requires a camera and a processor, but there is still more to it as shown in Figure 8.

The camera needs a lens to form the optical image. The lens must also provide an acceptable working distance (the distance from the lens to the scene), the right magnification so the scene fills the camera's view, and the ability to resolve detail. A light is also needed so the camera will have a sufficient amount of the right illumination to create a useable and reliable image.

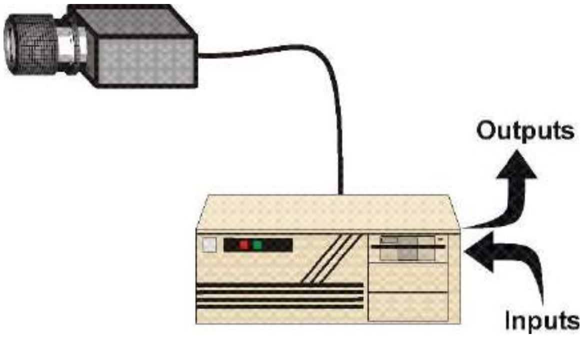


Figure 9 -- PC Based Vision System

There needs to be an interface between the camera and the processor. Software is needed to perform analysis needed by the application. Finally, the processor must support input and output needed by other connected equipment.

TYPES OF VISION SYSTEMS

Often the camera and processor are separate. This permits the camera to be smaller and the processor to have all the computing power and input-output capabilities required. It also allows one processor to serve two or more cameras. A common form of this type of vision system uses a PC as the processor (see Figure 9). The PC based vision system provides the most flexibility. There are many camera interfaces available, many different software packages available, and significant flexibility on the input-output configuration. However, this flexibility increases the engineering effort to address an application. Rugged PCs are available to make this configuration compatible with more demanding environments.

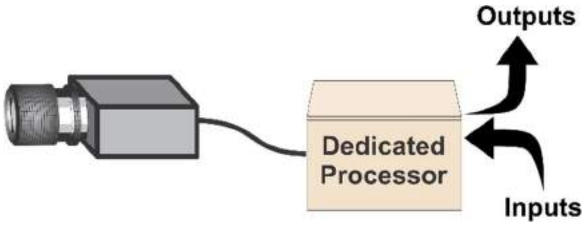


Figure 10 -- Vision System with Proprietary Processor

A similar configuration replaces the PC with a proprietary processor (see Figure 10). This processor comes bundled with proprietary software. It may use standard or proprietary camera interfaces and normally offers flexibility in input-output. This configuration is typically designed for the factory environment. The engineering effort to address an application is less than with a PC based vision system.

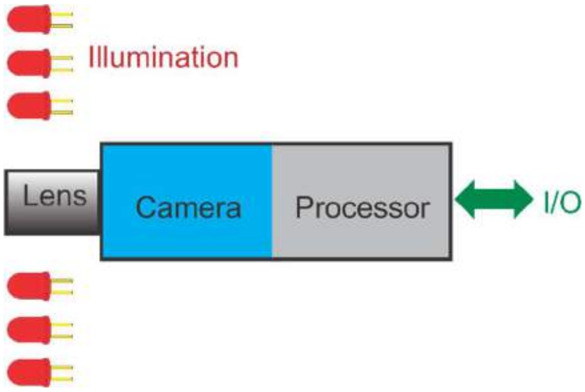


Figure 11 -- Smart Camera

There is a trend to combine the camera and processor into one compact device as shown in Figure 11 called a smart camera or smart sensor. This has become effective for single camera applications and for IoT (Internet-of-Things) devices. The processing and I/O are limited by the compact size. While the smart camera sometimes comes with built-in lighting, many times a different light source is needed. In most smart cameras, the software is incorporated into it to make these the easiest vision systems to apply. There are a few products, designed for the OEM who can afford more engineering, that do not come with software installed.

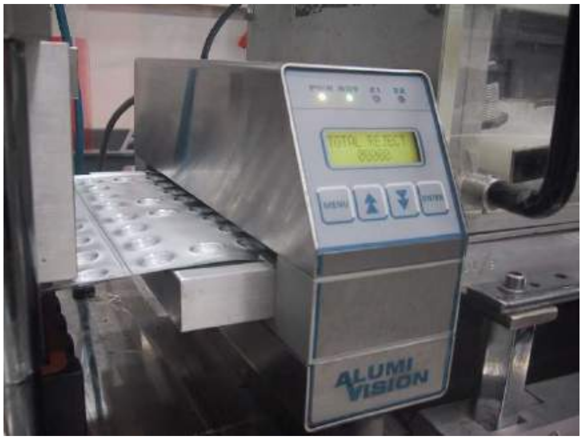


Figure 12 -- Application Specific Vision System

The final category is the application specific machine vision (ASMV) system. An example is shown in Figure 12. This is a vision system that has been designed for a specific application. Very little, if any, engineering, except for installation and product specific configuration, is required to make it ready to use. The higher cost of the ASMV is offset by the virtual elimination of application engineering. Also, more comprehensive field support is available than with other vision systems engineered for just one installation.

Figure 13 -- Embedded Vision System

The final category is the embedded vision system. There's not agreement on exactly what is an embedded vision system. For practical purposes, we'll say that it's a vision system designed to be incorporated into a final product and that has the camera and processor closely coupled such as shown in Figure 13. Mostly, these are intended for OEM integration into products.

THE CAMERA

Let’s look at the camera in a bit more detail. Figure 14 shows the elements of a camera.

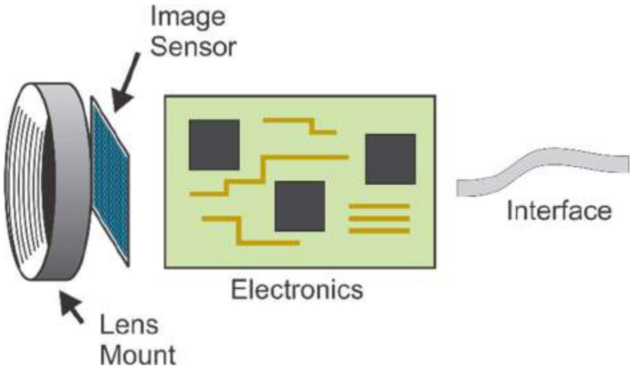


Figure 14 -- Camera Details

- The image sensor is the principal component of a camera and gives the camera most of its important characteristics.
- The electronics controls the image sensor’s timing, many of the image sensor’s functions like gain, and provides other features unique to the camera.
- The lens mount holds the lens that forms the optical image onto the image sensor.
- The interface connects the camera to the processor.

Image Sensor

The image sensor shown in Figure 15 consists of an array of sensing elements shown in the enlargement along with circuitry to read out the signal from the elements and turn them into digital numbers.

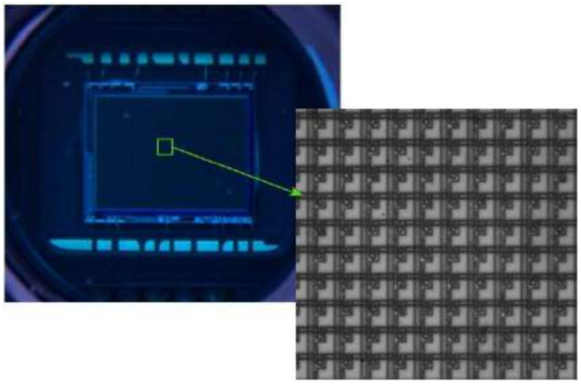


Figure 15 -- Image Sensor and Sensing Elements (Pixels)

When exposed to light the sensing elements convert incident photons into electrical charges (see Figure 16).

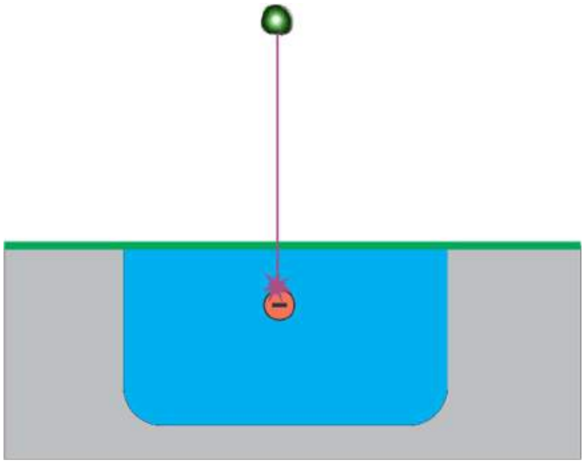


Figure 16 -- Photon Incident on a Photodetector

During exposure, these photogenerated electrons accumulate in the sensing element. When the sensing element is read out, the voltage from it is proportional to the number of photogenerated electrons. After readout, the sensing element is reset, eliminating all photogenerated electrons leaving it ready for the next exposure.

You could correctly deduce that the sensing element is actually a photon counter. This is almost correct except that it is an imperfect photon counter. It doesn't count every photon. The percentage of photons it counts is call its quantum efficiency, or QE for short. QE varies with the wavelength of the light as shown graphically in Figure 17

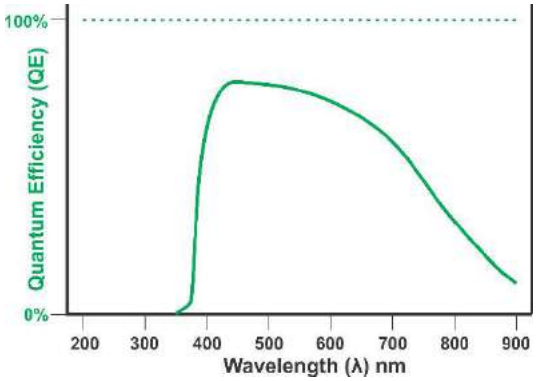


Figure 17 -- Quantum Efficiency (QE)

Color Imaging

There are two principal ways to make image sensors for color cameras. The first is to use three separate image sensors and optics to split the three color components, red, green, and blue into separate directions so that each of the three image sensors senses only one of the colors (see Figure 19). Each pixel has three values, one off of each of the three image sensors.

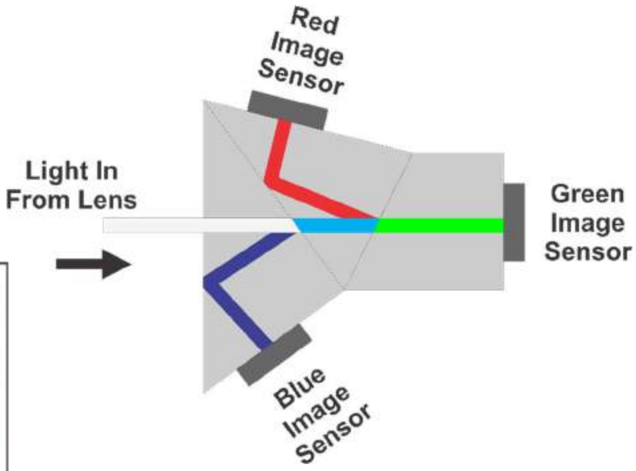


Figure 19 -- Three Chip Color Sensor

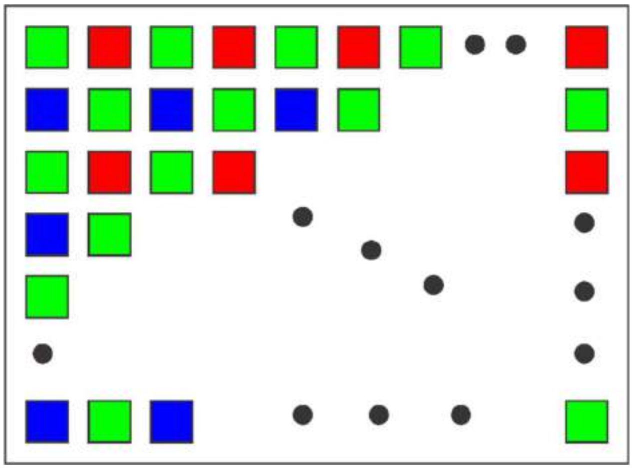


Figure 18 -- Single Chip Color Sensor with Bayer Filter Pattern

The other method is to put red, green, and blue color filters over each individual sensing element. The usual pattern for these filters is the Bayer pattern (see Figure 18). Each pixel still has three values: one that is sensed directly through its color filter with the other two color values being interpolated from neighboring pixels.

High-Dynamic Range

The noise level in cameras typically limits them to 8 or 10 bits of data per pixel. This corresponds to 256 or 1,024 levels of gray. For many applications, this is entirely adequate. However, some applications require perceiving detail in both the highlighted (bright) areas of an image as well as deep in shadows (dark) areas of an image.

The approach to solving this challenge is high-dynamic range (HDR) imaging. This involves taking two (or more) exposures, with different exposure times. The upper bits of pixels with the shorter exposure time will have details in the highlights, and the lower bits of pixels with the longer exposure time will have details in the shadows. A new pixel value is created by combining the bits from each exposure to give the higher dynamic range as shown in Figure 20.

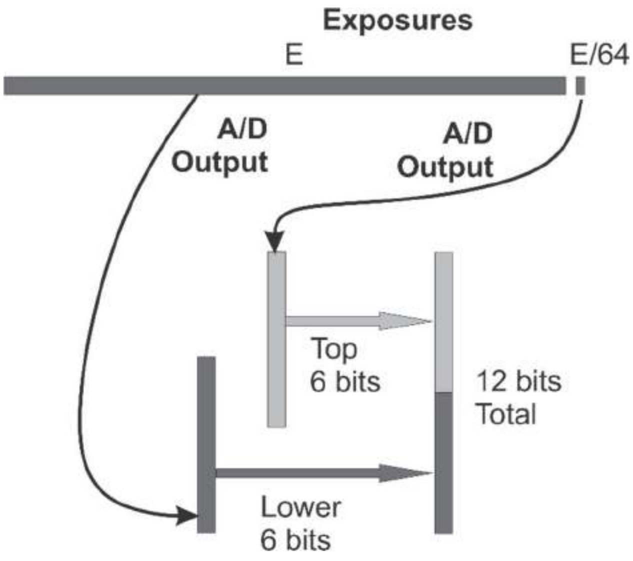


Figure 20 -- High-Dynamic Range

Figure 21 -- C-mount Lens Mount

Lens Mount

There are many different lens mounts used on machine vision cameras. The most common, though, is the C-mount. A lens mount is characterized by its thread (unless a bayonet mount) and its flange focal distance. The flange focal distance is the distance from the seating flange for the lens to the image sensor.

For the C-mount, the thread is 1 inch x 24 threads/inch. The flange focal distance is 17.52mm (0.69 inch). Another popular lens mount is the CS-mount. It has the same thread as the C-mount, but its flange focal distance is 12.52mm.

The lens must mount to the camera. Normally, the two components have the same mount. However, sometimes an adaptor is used when the lens and camera have different mounts.

Camera Interfaces

There are many camera interfaces. Six stand out as interfaces that are commonly used in machine vision:

- Camera Link
- GigE Vision
- USB3 Vision
- CoaXPress
- Camera Link HS
- MIPI

Most of these standards have options, some many options. That makes a description of each interface beyond the scope of this whitepaper. The interfaces vary in the following important characteristics:

- Bandwidth/speed – how fast the image data can be transmitted from the camera to the processor.
- Latency – the time between when the image data is started to when it is complete. The variation in latency is often the most important factor
- Data reliability – if the data is corrupted during transmission, does the interface detect this and can it correct the corruption?
- Cable length – how long a cable can be used and still have adequate bandwidth and data reliability.
- Cable/connector – what kind of cable and what kind of connector is used. Many of the standards provide for different cables and different connectors. Most interface standards also address using fiber optic cables.
- Frame grabber – does the interface require the use of a special adaptor card called a frame grabber.
- Power – can the camera's power be supplied over the same cable as used for the data.

Another important interface standard is GenICam. GenICam is not a physical standard; it has nothing to do with cables, data rates, etc. It is a standard about how a camera informs the processor its connected to about the camera's characteristics. Under GenICam, the camera hosts a XML file that can be read by the processor. This file describes the camera and all its settings in detail. The software in a processor can interrogate this file to learn everything it needs to interface and control the camera. Different cameras, even from the same manufacturer, have different ways of being controlled. GenICam makes this transparent to the vision system developer. Allowing the developer to program an application without concern for what camera is connected.

THE LENS

The lens is a critical component of a vision system. It, along with the camera's image sensor, determines the working distance and the field-of-view that encompasses the scene. The lens with its aperture also determines how much light energy reaches onto the image sensor, the depth-of-field or range of distances that are in focus, and the detail that is resolved in the optical image that is projected onto the image sensor.

Most experienced machine vision engineers have learned how to select lenses for most applications. For people new to machine vision or for unusually difficult applications, the use of an experienced optics engineer may be needed.

Figure 22 -- Ordinary or Entrocentric Lens

There are three basic types of lenses: the ordinary, or entocentric lens(Figure 22), the macro lens(Figure 23), and the telecentric lens (Figure 25). Most applications are solved with the use of the entocentric lens. These lenses can be focused over a range of working distance extending to infinity and usually have an adjustable aperture.

Because entocentric lenses do not focus closely to achieve high magnifications, macro lenses are designed to fill this requirement. Magnification is the ratio of the size of the image divided by the size of the imaged scene. Macro lenses generally work in the magnification range of .05x to 10x. Where higher magnification is needed, microscopes are usually the better choice.

Figure 23 -- Macro Lens

Figure 25 -- Telecentric Lens

Telecentric lenses are specialty lenses that find frequent use in machine vision. Unlike entrocentric lenses where the magnification changes with a change in working distance, within its depth-of-field, the telecentric lens' magnification is essentially constant. This means that measurements are virtually unaffected by modest changes in a part's height and that there is no perspective distortion.

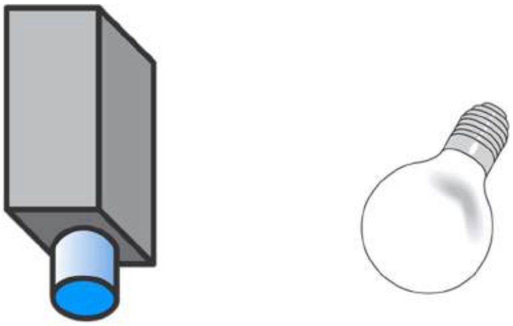


Figure 24 -- Front Lighting

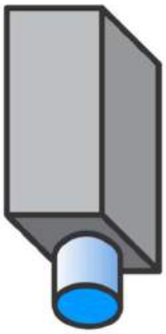
While zoom lenses would seem to be ideal for machine vision giving the ability to set the field-of-view to whatever size is needed, they are used only sparingly. Because there is one more adjustment, it is difficult to ensure that the lens is set to the correct magnification, and calibration may be compromised. Also, zoom lenses tend to be larger, heavier, and more expensive than entrocentric lenses.

LIGHTING

Good lighting is essential for a machine vision system that must work reliably. That is why industrial machine vision systems always incorporate engineered lighting and minimize the influence of ambient light.

Some of the more recent developments, like autonomous driven vehicles, are constrained to use ambient light almost exclusively.

For engineered lighting, there are several ways to differentiate the lighting techniques. The first is front lighting (Figure 24) or back lighting (Figure 26). In front lighting, the light source is on the same side of the scene as the camera. In back lighting, the light is on the opposite side of the scene from the camera.



Another differentiator in lighting is brightfield and darkfield (see Figure 27). Brightfield light is any light that is expected to be reflected (off a mirror) or transmitted directly into the camera's view. Darkfield light is any light that is positioned so that it will not transmit or be reflected (off a mirror) into the camera's view.

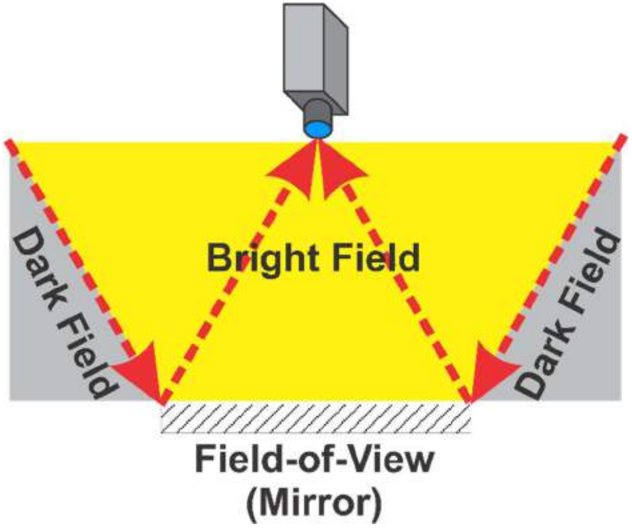


Figure 27 -- Bright Field and Dark Field Illumination

As shown in Figure 28, another differentiation in lighting technique is the degree to which a light is directed or diffuse. Directed light illuminates every point on the scene with a single light ray, or more commonly with a group of rays that are spread in a very narrow range of angles.

Figure 26 -- Back Lighting

Diffuse illumination illuminates every point on the scene with a bundle of rays coming from a wide range of angles.

Directed illumination tends to cast strong shadows and make texture stand out in the image. Diffuse illumination casts faint shadows or no shadows and eliminates the shading that makes texture evident in the image.

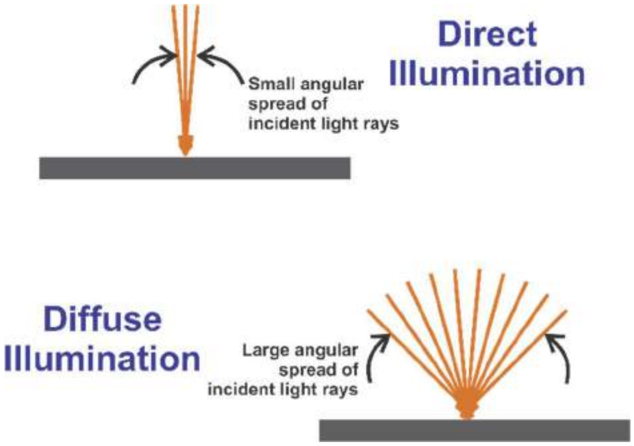


Figure 28 -- Directed and Diffuse Illumination

A final characteristic of the lighting design is color or the spectrum of light used. Most machine vision systems work in the range of visible light. For monochrome imaging, the light is usually a narrow band of wavelengths either red, or green, or blue. For color imaging, white light is needed. There are a growing number of applications using infrared wavelengths. Since humans cannot see in this region of the spectrum, some experimentation is needed to know how it will work. Ultraviolet light is also used, usually to make parts fluoresce.

What this means is that the vision system engineer has a wide variety of light sources covering a wide range of sizes, a wide range of colors, and providing a wide range of illumination characteristics. To know how to select the right light source takes study and experience. People starting out in machine vision would do well to work with someone more experienced when picking out the light source.

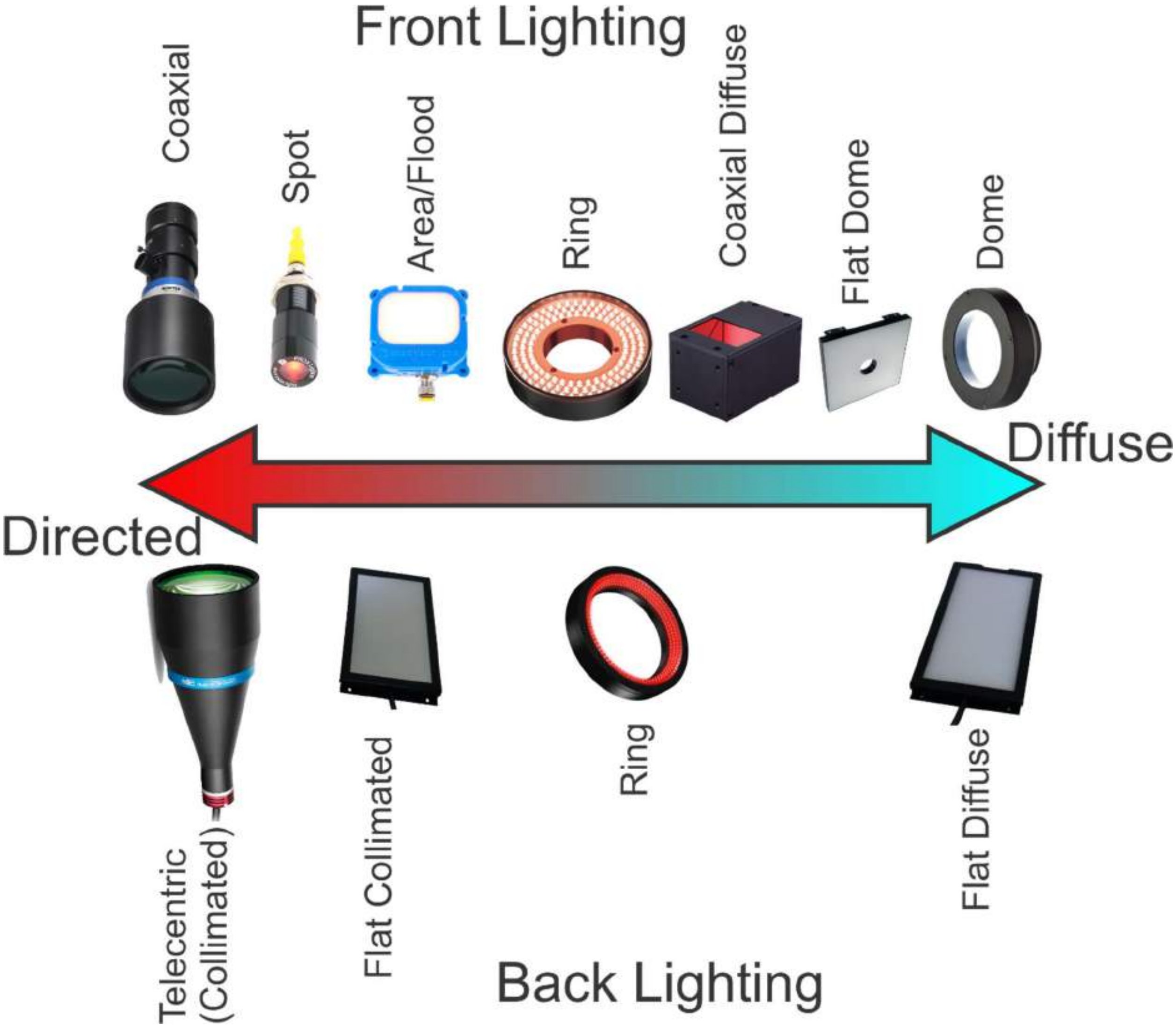


Figure 29 -- Array of Light Sources

IMAGE PROCESSING

The big question in image processing is whether to use algorithms, tools, or deep learning. Let's first look at the image processing process which is actually a taxonomy for algorithms to perform image processing.

IMAGE PROCESSING PROCESS

Figure 30 -- The Machine Vision Image Processing Process

There are basically four steps in image processing bounded at the beginning by a process to bring in the image and at the end by a process to deliver the results. The process, illustrated in Figure 30, shows the four steps: preprocessing, segmentation, feature extraction, and interpretation. This figure is somewhat simplified. Not all applications require all four steps. Some applications do not need preprocessing and some applications do not need segmentation. Any one of these steps may require one or more algorithms. Frequently, there are feedback loops where preliminary analysis selects a particular sequence of steps depending on the detected image characteristics and contents.

TRADITIONAL ALGORITHMS

Preprocessing

Preprocessing algorithms have an image as an input and provide a modified image as an output. For example, the algorithm might filter out noise (Figure 31), or it might extract edges (Figure 32). There are an enormous number of preprocessing operators. A very experienced image processing programmer will know which few are good candidates for any application. Still, it takes a lot of experimentation to pick the right operator and make the fine adjustments to have it work reliably over all conditions.

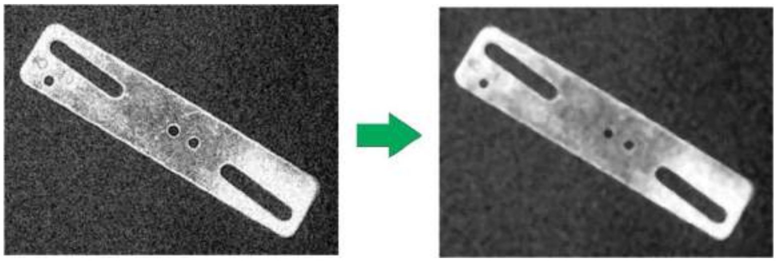


Figure 31 -- Preprocessing: Low-Pass Filter

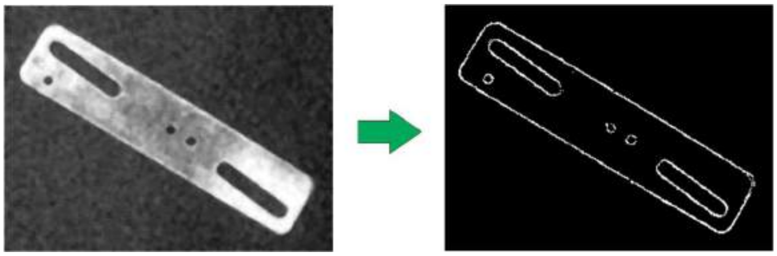


Figure 32 -- Preprocessing: Edge Detection

Preprocessing is not needed by all machine vision applications.

Segmentation

It is often necessary to segment an image: identify the different regions within an image. Again, there are a variety of techniques available to the programmer. See Figure 34 and Figure 33 for examples. Usually, the requirements of the application make selecting the right one obvious to a skilled image processing programmer, and, again, time and experimentation are needed to tune the segmentation for optimum results on the application's images.

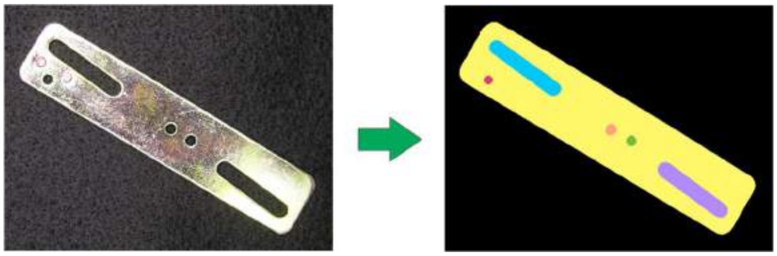


Figure 34 -- Segmentation: Region Labeling

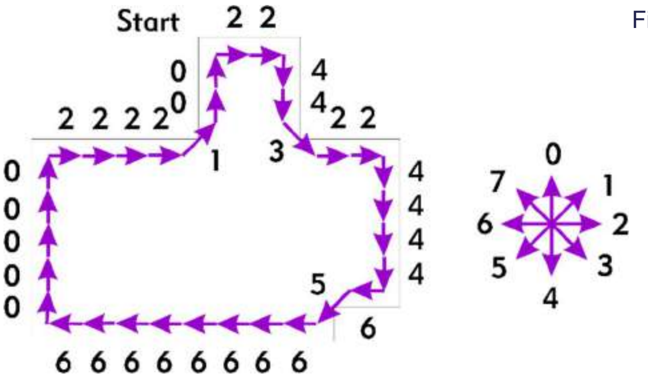


Figure 33 -- Segmentation: Boundary Tracing

Like preprocessing, segmentation is not needed by all applications.

Feature Extraction

The definition of machine vision makes it apparent that getting data from images is essential. The features extracted from the images is that data. Every machine vision process must extract features from the image. See Figure 35 and Figure 36 for examples.

There are literally hundreds of features that may be extracted from the parts in an image. Typically, only a handful of features are needed. Selecting the features needed is usually evident from the requirements for the vision system.

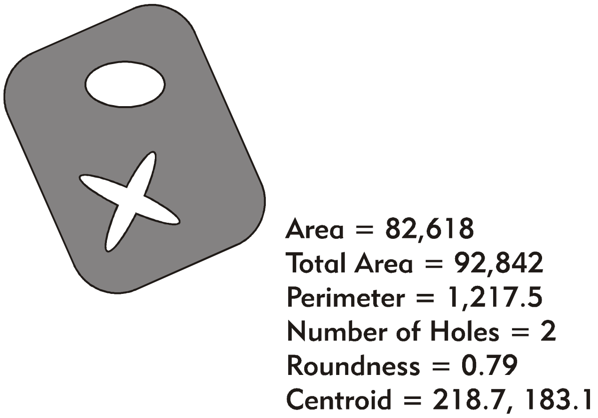


Figure 35 -- Features of a Part

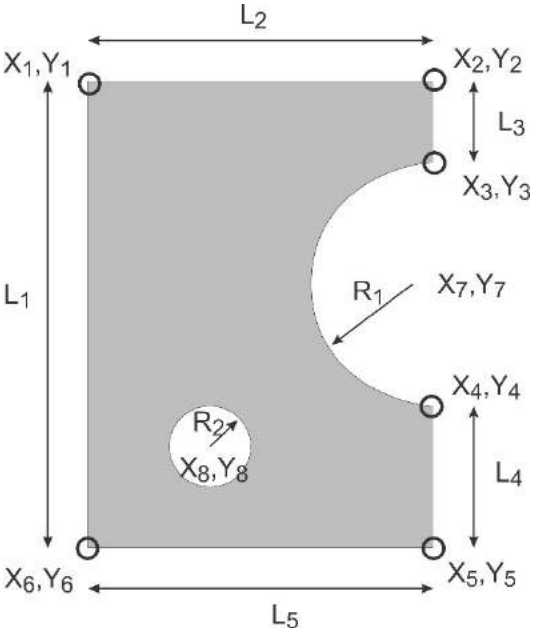


Figure 36 -- Dimensional Features

The definition of machine vision makes it apparent that getting data from images is essential. The features extracted from the images is that data.

Interpretation

Having features is not the end. There needs to be interpretation. Is the part good or bad? Where send the robot to pick up the part? Which bin should be part be put into?

The classic method of interpretation is the decision tree (Figure 37). It's made up of "if-then-else" statements familiar to anyone who has done even a small amount of programming.

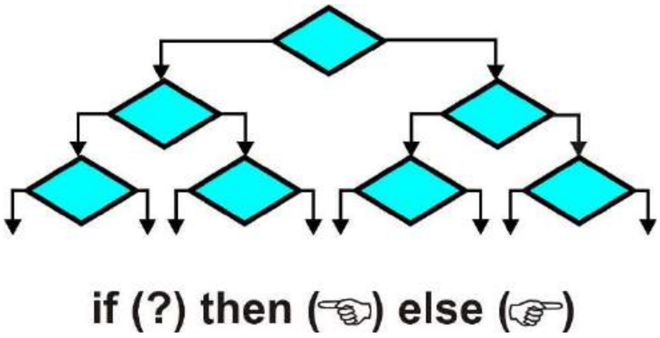


Figure 37 -- Decision Tree

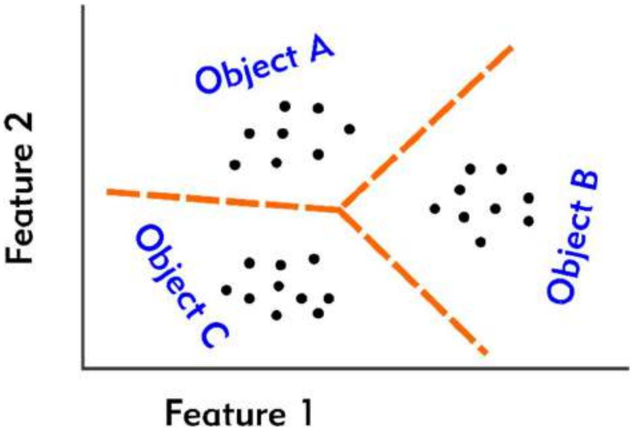


Figure 38 -- Linear Discriminator

Other machine learning techniques also have a place such as linear discriminators (Figure 38).

IMAGE ANALYSIS TOOLS

Most machine vision software packages offer tools for performing tasks that eliminate the need for the machine vision programmer to sort through and experiment with hundreds of algorithms. These tools address specific functions such as:

- Object recognition, location, and orientation
- Measurement (e.g., distance, width, diameter)
- Defect detection
- Seam tracking
- OCR (optical character recognition)
- OCV (optical character verification)
- Code reading (bar codes and 2D codes)

DEEP LEARNING

The latest major trend in machine vision image analysis is deep learning. Deep learning uses a neural network with images as its input and many internal layers. Internal layers are not visible to the user or (usually) not even the vision system engineer. A suite of training images are used to train the network. During each training pass, the coefficients in every layer are adjusted by some small amount depending on the error at the output.

The most common network approach is the convolutional neural network (CNN). This approach has been described as programming with data rather than with computer code.

While deep learning is solving some very difficult problems in machine vision, it's not a panacea. Experts agree that if an application can be solved with tools and algorithms without undue effort, that route will be easier than using deep learning.

While deep learning is "programming with data", it requires a large collection of correctly labeled data to train a network to perform reliably. The cost of data acquisition and labeling is not at all trivial and takes much more time than training the network.

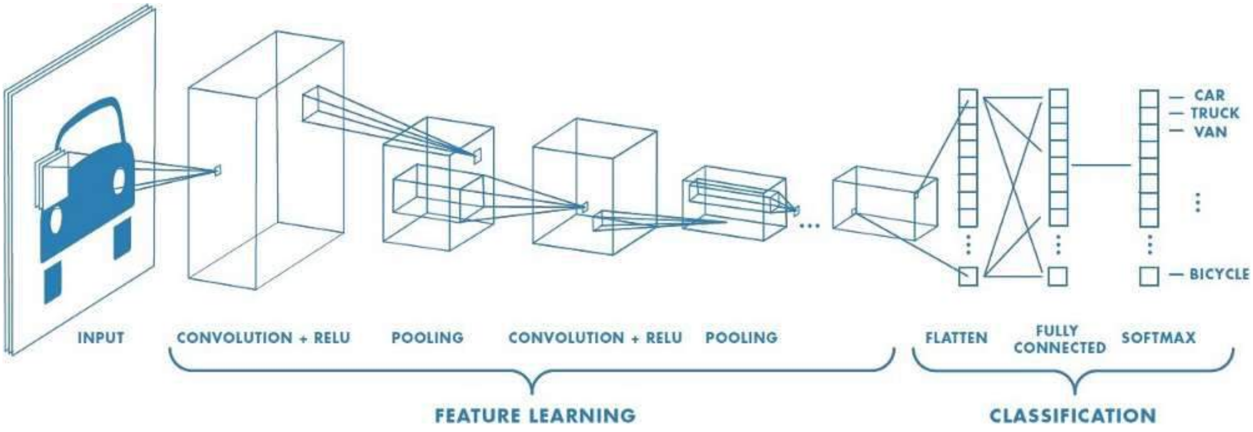


Figure 39 -- Convolutional Neural Network

GLOSSARY

acquisition – The process of creating and making the electronic or digitized image available for processing.

analysis – The activity of processing an image to extract desired data or actions needed.

aperture – The effective diameter of a lens that controls the amount of light reaching the image. It is usually controlled by an adjustable iris.

back lighting – Any illumination method in which the light source and camera are on opposite sides of the illuminated scene.

Bayer pattern -- A repetitive pattern of alternating rows of colored filters placed over individual sensing elements in an image sensor. One row has alternating red and green filters and the other row has alternating green and blue filters. The rows are alternated down the image sensor.

binarization – The process of converting a gray-scale image into a binary image. (See also thresholding.)

binary – 1) A signal or item of data that has only two states: on or off, true or false, one or zero, or high or low. 2) An image with pixels that are only either light or dark.

boundary tracing – A process that segments two regions in an image by tracing along the interface between the two regions.

bright field – In illumination, lighting that either shines directly into the camera's lens or would be reflected into the lens if the scene was a mirror.

camera – In machine vision, the component that contains the image sensor and interfaces to a processor. It also usually has a mount for attaching the lens.

CNN – An acronym for convolutional neural network.

convolution – A preprocessing operation that operates mathematically on successive groups of pixels to create a new pixel value for the resulting image.

convolutional neural network -- A neural network designed for deep learning that takes an image as its input. The first few layers of the network are designed to perform convolutions.

dark field – An illumination technique where the illumination is supplied at a grazing angle to the surface.

decision tree – A network of decisions expanding from a single input to two or more outputs. Usually implemented in software with a series of nested "if ... then ... else" statements.

deep learning – Any machine learning technique, often a convolutional neural network, that has as its inputs image data, with many hidden layers, and provides useful outputs that are learned from example images.

depth-of-field – The in-focus range (for objects) of an imaging system.

diffuse illumination – Illumination in which the light reaching any point in the scene comes from a wide range of angles. It produces soft or almost no shadows.

digital image – An image which is digitized into discrete samples (pixels) each quantized to one of a possible number of discrete amplitudes.

directed illumination – Light that reaches the scene through a narrow range of angles such that it casts strong shadows.

edge detection – A process that identifies edges in an image and excludes image regions that are not edges.

engineered lighting – Illumination that has been purposefully implemented to make a machine vision system work reliably.

entocentric – the type of lens used most commonly in machine vision. It is typically focusable from infinity to some close distance and has an adjustable iris.

feature extraction – The process of obtaining attribute data from an image.

field-of-view – The region of the scene that is imaged by an imaging system.

flange focal distance – The distance from the lens seating flange to the image plane.

frame grabber – A device that interfaces with a camera and stores the digitized image in a processor’s memory.

front lighting – Illumination arranged so that light reaching the image sensor is reflected off the objects in the scene. The light source and camera are on the same side of the object.

gray-scale – The range of possible values of light magnitude representable in a pixel or image.

HDR – Acronym for high-dynamic range.

high-dynamic range – An imaging technique to improve detail in highlights and shadows of an image by increasing the apparent dynamic range. It uses two or more exposures and compiles a pixel value by combining portions of the digitized amplitude of each exposure.

image – A representation of a scene which contains one or more physical objects.

image resolution – The number of rows and columns of pixels in an image.

image sensor – A device which can convert an optical image into an electrical signal. Commonly, image sensors are semiconductor devices consisting of a regular array of photosensors together with circuits to read out the light sensed by each sensing element.

interface – The hardware and protocols for transmitting or exchanging image data.

interpretation – Results of the image processing that satisfy the output required for the system.

iris – An adjustable aperture built into a lens to permit control of the amount of light passing through the lens.

lens – One or more transparent, refractive optical components with one or more curved surfaces (that cause incident light rays to converge to create an optical image of a scene.

lens mount – A structure to support and maintain alignment of a lens assembly.

linear classifier – A classifier, usually able to learn from examples, that make a decision based on a linear combination of the characteristics.

low-pass filter -- A filter which passes lower frequencies and suppresses higher frequencies. The effect of a low-pass filter on a digitized image is to soften (blur) the edges and reduce noise.

machine vision – The automatic acquisition and analysis of images to obtain desired data for controlling an activity or process.

macro lens – An imaging lens designed to give best performance when used at magnifications around 1.

magnification – The image size divided by the object size.

monochrome – Black and white with shades of gray. Excludes color.

object recognition – A match between a description derived from an image and a description obtained from a stored model.

OCR – Acronym for optical character recognition.

ODV – Acronym for optical character verification.

optical character recognition – Recognition of each individual character in a string by a vision system.

optical character verification – Using a vision system to check that each character in a string is the correct character and that it is legible and meets quality standards.

optical image – Light energy brought to a focus so that it faithfully represents the scene imaged.

pixel – A contraction of the term picture element. 1) The individual elements in a digitized image array. 2) In a solid-state image sensor, a synonym for sensing element.

point cloud – Three-dimensional image data where each sensed point has an X, Y, and Z coordinate, but the distance between points is not necessarily constant.

preprocessing – Operations used in machine vision to enhance an image. This consists of image processing which transforms one digitized image into another digitized image.

QE – Acronym for quantum efficiency.

quantum efficiency – For image sensors, the ratio of the number of photogenerated electrons to the number of incident photons.

region labeling – In image processing, a segmentation process that labels each region of similar pixels in order to distinguish it from every other region.

scene – The area imaged by a camera. Often used interchangeably with object plane and field-of-view.

segmentation – In image analysis, the process of identifying individual contiguous regions within an image usually corresponding to surfaces, objects, or entities in the scene.

sensing element – A discrete device on an image sensor for sensing light energy falling on a defined small area. Also, called pixel.

telecentric lens – A lens designed with a small aperture at one foci so that it images all points on the scene with parallel paths. It virtually eliminates parallax and perspective distortion as well as magnification variation with a change in working distance.

three chip color sensor – A color image sensing technique that uses three monochrome image sensors and a special prism that splits the light energy in three directions based on its wavelength. The three directions typically correspond to red, green, and blue colors.

threshold – 1) A value used in converting a gray-scale image to a binary image. All pixels with values equal to or greater than the threshold are set light; all other pixels are set dark. 2) Any single valued decision criteria.

thresholding – Separating regions of an image based on pixel values above or below a threshold value.

working distance – It is the distance between the scene and front of the lens.

zoom lens – A lens of continuously variable focal length; the focal plane (image position) remains in a fixed position as the focal length is varied.

ACKNOWLEDGEMENTS

Figure 7 – courtesy of Mathworks

Figure 12 – Courtesy of Electronic Development Solutions

Figure 13 – Courtesy of Teknique

Camera Link is a registered trademark of the Association for Advancing Automation (A3)

GigE Vision is a registered trademark of the Association for Advancing Automation (A3)

USB3 Vision is a registered trademark of the Association for Advancing Automation (A3)

CoaXPress is a registered trademark of the Japan Industrial Imaging Association (JIIA)

Camera Link HS is a registered trademark of the Association for Advancing Automation (A3)

MIPI is a registered trademark of the MIPI Alliance, Inc.

GenICam is a registered trademark of the European Machine Vision Association (EMVA)

Figure 23 courtesy of Computar

Figure 25 courtesy of Moritex

Figure 29 -- Advanced Illumination, Opto Engineering, CCS America, Smart Vision Lights, Moritex North America, Inc., Omron Corp. Metaphase Technologies Inc.

Figure 39 – from Towards Data Science