



Machine Vision Introduction

SICK|IVP

© SICK IVP
Version 2.2, December 2006
All rights reserved
Subject to change without prior notice

Contents

1	Introduction.....	7
1.1	Objective	7
1.2	Application Types	8
1.2.1	Locate.....	8
1.2.2	Measure	8
1.2.3	Inspect.....	8
1.2.4	Identify.....	8
1.3	Branch Types	9
1.4	Camera Types	9
1.4.1	Vision Sensors	9
1.4.2	Smart Cameras.....	10
1.4.3	PC-based Systems.....	11
2	Imaging.....	12
2.1	Basic Camera Concepts.....	12
2.1.1	Digital Imaging.....	12
2.1.2	Lenses and Focal Length.....	13
2.1.3	Field of View in 2D.....	14
2.1.4	Aperture and F-stop.....	14
2.1.5	Depth of Field	15
2.2	Basic Image Concepts	16
2.2.1	Pixels and Resolution.....	16
2.2.2	Intensity.....	17
2.2.3	Exposure	18
2.2.4	Gain	19
2.2.5	Contrast and Histogram.....	19
3	Illumination.....	21
3.1	Illumination Principles	21
3.1.1	Light and Color	21
3.1.2	Reflection, Absorption, and Transmission	22
3.2	Lighting Types.....	23
3.2.1	Ring Light	23
3.2.2	Spot Light.....	23
3.2.3	Backlight	24
3.2.4	Darkfield.....	24
3.2.5	On-Axis Light	25
3.2.6	Dome Light.....	26
3.2.7	Laser Line	26
3.3	Lighting Variants and Accessories.....	27
3.3.1	Strobe or Constant Light.....	27
3.3.2	Diffusor Plate.....	27
3.3.3	LED Color	27
3.3.4	Optical Filters.....	28
3.4	Safety and Eye Protection	29
3.4.1	Laser Safety	29
3.4.2	LEDs	30
3.4.3	Protective Eyewear.....	30
4	Laser Triangulation	31
4.1	Field of View in 3D.....	31

4.2	3D Image and Coordinate System.....	32
4.3	Scanning Speed.....	32
4.4	Occlusion and Missing Data.....	33
4.5	System Components.....	34
4.6	Ambient Light Robustness.....	34
5	Processing and Analysis.....	35
5.1	Region of Interest.....	35
5.2	Pixel Counting	35
5.3	Digital Filters and Operators.....	36
5.4	Thresholds.....	37
5.5	Edge Finding	38
5.6	Blob Analysis.....	38
5.7	Pattern Matching.....	39
5.8	Coordinate Transformation and Calibration	39
5.9	Code Reading	40
5.9.1	Barcode	40
5.9.2	Matrix Code.....	40
5.10	Text Verification and Reading	40
5.10.1	Optical Character Verification: OCV	40
5.10.2	Optical Character Recognition: OCR.....	41
5.11	Cycle Time	42
5.12	Camera Programming.....	42
6	Communication.....	44
6.1	Digital I/O	44
6.2	Serial Communication.....	44
6.3	Protocols	44
6.4	Networks	45
6.4.1	Ethernet.....	45
6.4.2	LAN and WAN.....	45
7	Vision Solution Principles.....	46
7.1	Standard Sensors.....	46
7.2	Vision Qualifier.....	46
7.2.1	Investment Incentive.....	46
7.2.2	Application Solvability	46
7.3	Vision Project Parts	47
7.3.1	Feasibility Study.....	47
7.3.2	Investment	47
7.3.3	Implementation	47
7.3.4	Commissioning and Acceptance Testing	47
7.4	Application Solving Method	48
7.4.1	Defining the Task	48
7.4.2	Choosing Hardware	48
7.4.3	Choosing Image Processing Tools	48
7.4.4	Defining a Result Output	48
7.4.5	Testing the Application	48
7.5	Challenges	49
7.5.1	Defining Requirements	49
7.5.2	Performance	49
7.5.3	System Flexibility	49
7.5.4	Object Presentation Repeatability	49
7.5.5	Mechanics and Environment.....	49

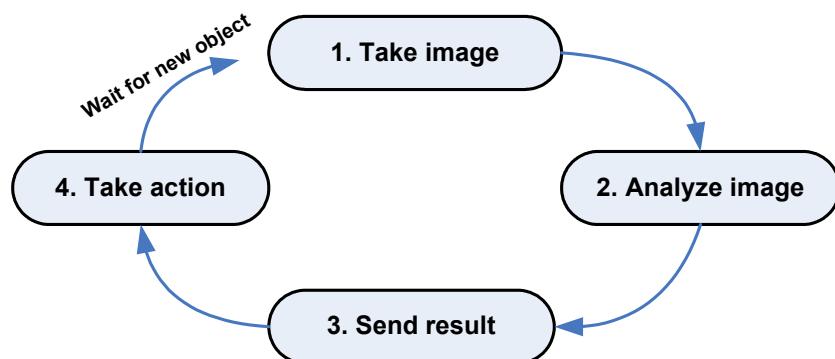
8	Appendix	50
A	Lens Selection	50
B	Lighting Selection.....	52
C	Resolution, Repeatability, and Accuracy.....	53
D	Motion Blur Calculation	54
E	IP Classification	54
F	Ethernet LAN Communication.....	55

1 Introduction

Machine vision is the technology to replace or complement manual inspections and measurements with digital cameras and image processing. The technology is used in a variety of different industries to automate the production, increase production speed and yield, and to improve product quality.

Machine vision in operation can be described by a four-step flow:

1. **Imaging:** Take an image.
2. **Processing and analysis:** Analyze the image to obtain a result.
3. **Communication:** Send the result to the system in control of the process.
4. **Action:** Take action depending on the vision system's result.



This introductory text covers basic theoretical topics that are useful in the practical work with machine vision, either if your profession is in sales or in engineering. The level is set for the beginner and no special knowledge is required, however a general technical orientation is essential. The contents are chosen with SICK IVP's cameras in mind, but focus is on understanding terminology and concepts rather than specific products.

The contents are divided into eight chapters:

1. Introduction (this chapter)
2. Imaging
3. Illumination
4. Laser Triangulation
5. Processing and Analysis
6. Communication
7. Vision Solution Principles
8. Appendix

The appendix contains some useful but more technical issues needed for a deeper understanding of the subject.

1.1 Objective

The objective is that you, after reading this document:

1. Understand basic machine vision terminology.
2. Are aware of some possibilities and limitations of machine vision.
3. Have enough theoretical understanding to begin practical work with machine vision.

1.2 Application Types

Machine vision applications can be divided into four types from a technical point of view: **Locate, measure, inspect, and identify.**

1.2.1 Locate

In locating applications, the purpose of the vision system is to find the object and report its **position** and **orientation**.

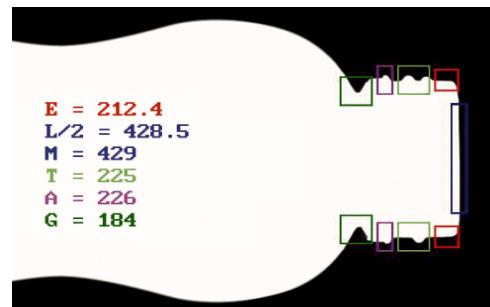
In robot bin picking applications the camera finds a reference coordinate on the object, for example center of gravity or a corner, and then sends the information to a robot which picks up the object.



1.2.2 Measure

In measurement applications the purpose of the vision system is to measure **physical dimensions** of the object. Examples of physical dimensions are distance, diameter, curvature, area, height, and volume.

In the example to the right, a camera measures multiple diameters of a bottleneck.



1.2.3 Inspect

In inspection applications the purpose of the vision system is to validate certain features, for example **presence** or **absence** of a correct label on a bottle, screws in an assembly, chocolates in a box, or defects.

In the example to the right, a camera inspects brake pads for defects.

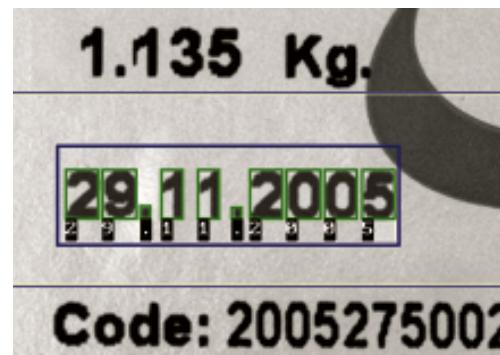


1.2.4 Identify

In an identification application the vision system reads various **codes** and alphanumeric **characters** (text and numbers).

In the example to the right, a camera reads the best before date on a food package.

Examples of codes that can be read simultaneously on the same package are barcodes and matrix codes.



1.3 Branch Types

Machine vision applications can also be categorized according to **branch** type, for example:

- Automotive
- Electronics
- Food
- Logistics
- Manufacturing
- Robotics
- Packaging
- Pharmaceutical
- Steel and mining
- Wood.

The branch categories often overlap, for example when a vision-guided robot (robotics) is used to improve the quality in car production (automotive).

1.4 Camera Types

Cameras used for machine vision are categorized into **vision sensors**, **smart cameras**, and **PC-based systems**. All camera types are digital, as opposed to analog cameras in traditional photography. Vision sensors and smart cameras analyze the image and produce a result by themselves, whereas a PC-based system needs an external computer to produce a result.

1.4.1 Vision Sensors

A **vision sensor** is a specialized vision system that is configured to perform a specific task, unlike general camera systems that have more flexible configuration software.

Thanks to the specific functionality of the vision sensor, the setup time is short relative to other vision systems.

Example

The CVS product range includes vision sensors for color sorting, contour verification, and text reading functionality.

For example, the vision sensors are used to inspect lid color on food packages and to verify best before dates on bottles.



Lid color verification on food packages.



Best before date inspection on bottles.

1.4.2 Smart Cameras

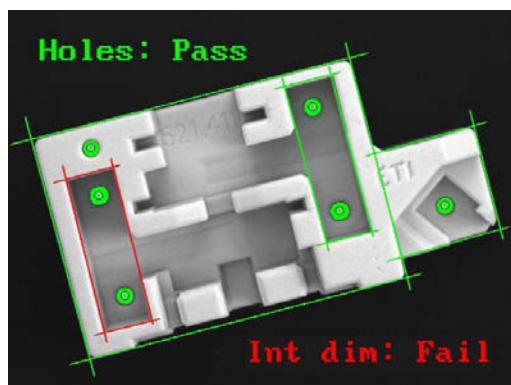
A **smart camera** is a camera with a built-in image analysis unit that allows the camera to operate **stand alone** without a PC. The flexible built-in image analysis functionality provides inspection possibilities in a vast range of applications.

Smart cameras are very common in 2D machine vision. SICK IVP also produces a smart camera for 3D analysis.

Example: 2D Smart

The IVC-2D (Industrial Vision Camera) is a stand-alone vision system for 2D analysis.

For example, the system can detect the correct label and its position on a whisky cap. A faulty pattern or a misalignment is reported as a fail.



Measurement of ceramic part dimensions.

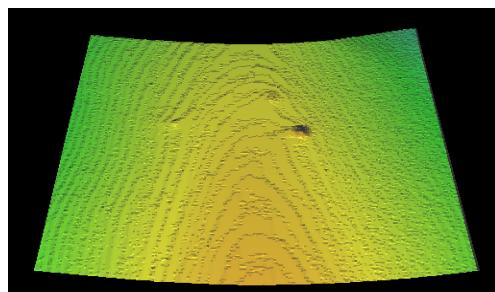


Misaligned label on the cap.

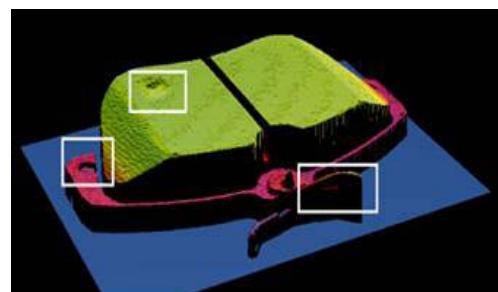
Example: 3D Smart

The IVC-3D is a stand-alone vision system for 3D analysis. It scans calibrated 3D data in mm, analyzes the image, and outputs the result.

For example, the system can detect surface defects, measure height and volume, and inspect shape.



Scanned wood surface with defects.



Brake pad (automotive) with defects.

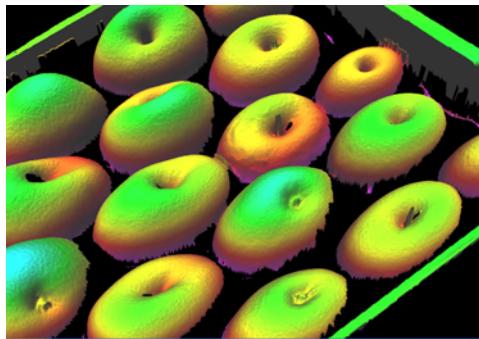
1.4.3 PC-based Systems

In a **PC-based system** the camera captures the image and transfers it to the PC for processing and analysis. Because of the large amounts of data that need to be transferred to the PC, these cameras are also referred to as **streaming devices**.

Example: 3D Camera

The Ruler collects calibrated 3D-shape data in mm and sends the image to a PC for analysis.

For example, it detects the presence of apples in a fruit box and measures log dimensions to optimize board cutting in sawmills.



Volume measurement and presence detection of apples in a box.



Log scanning for knot detection and board cutting optimization.

Example: MultiScan Camera

The Ranger has a unique **MultiScan** functionality that can perform multiple scans simultaneously in one camera unit, for example generating a 2D gray scale and a 3D image in one scan. MultiScan is accomplished by simultaneous line scanning on different parts of the object, where each part is illuminated in a special way.

Example

The Ranger C55 (MultiScan) scans three different kinds of images of a CD simultaneously:

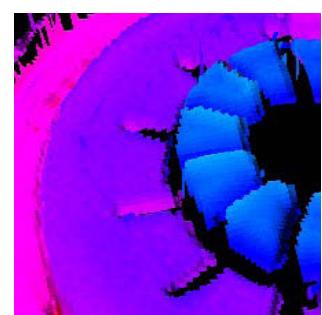
1. Gray scale for print verification
2. Gloss for crack detection
3. 3D for shape verification.



Gray scale.



Gloss.



3D.

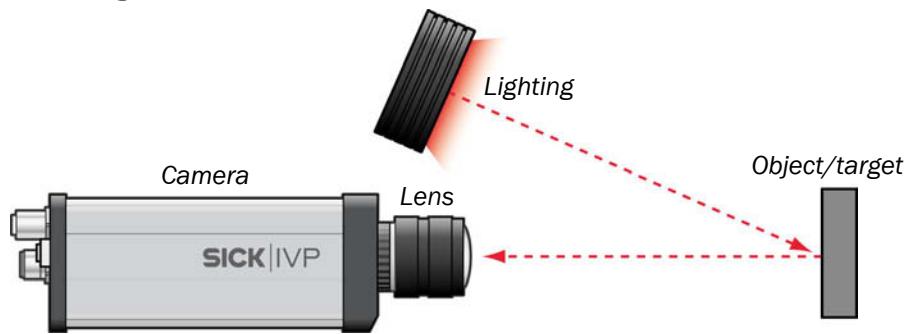
2 Imaging

The term **imaging** defines the act of creating an **image**. Imaging has several technical names: **Acquiring**, **capturing**, or **grabbing**. To **grab** a high-quality image is the number one goal for a successful vision application.

This chapter covers the most basic concepts that are essential to understand when learning how to grab images.

2.1 Basic Camera Concepts

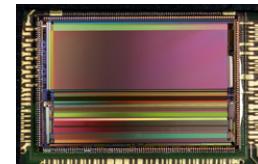
A simplified camera setup consists of **camera**, **lens**, **lighting**, and **object**. The lighting illuminates the object and the reflected light is seen by the camera. The object is often referred to as **target**.



2.1.1 Digital Imaging

In the digital camera, a **sensor** chip is used to grab a **digital image**, instead of using a photographic film as in traditional photography.

On the sensor there is an **array** of light-sensitive **pixels**. The sensor is also referred to as **imager**.



Sensor chip with an array of light-sensitive pixels.



There are two technologies used for digital image sensors:

1. **CCD** (Charge-Coupled Device)
2. **CMOS** (Complementary Metal Oxide Semiconductor).

Each type has its technical pros and cons. The difference of the technologies is beyond the scope of this introductory text.

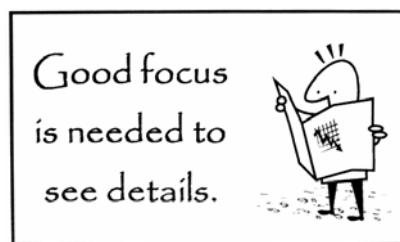
In PC-based systems, a **frame grabber** takes the raw image data into a format that is suitable for the **image analysis software**.

A **line scan camera** is a special case of the above where the sensor has only one pixel row. It captures one line at a time, which can either be analyzed by itself or several lines can be put together to form a complete image.

2.1.2 Lenses and Focal Length

The **lens** focuses the light that enters the camera in a way that creates a sharp image. Another word for lens is **objective**.

An image in **focus** means that the object edges appear sharp. If the object is **out of focus**, the image becomes **blurred**. Lenses for photography often have auto-focus, whereas lenses for machine vision either have a fixed focus or manually adjustable focus.



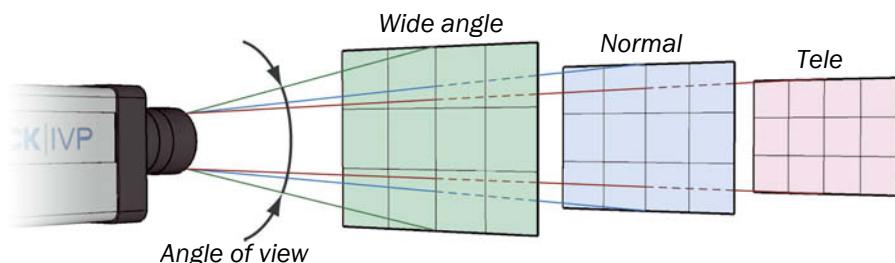
Focused or sharp image.



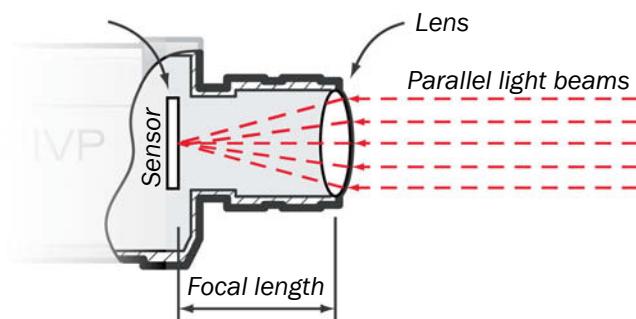
Unfocused or blurred image.

The main differences between lens types are their **angle of view** and **focal length**. The two terms are essentially different ways of describing the same thing.

The angle of view determines how much of the visual scene the camera sees. A **wide angle** lens sees a larger part of the scene, whereas the small angle of view of a **tele** lens allows seeing details from longer distances.



The focal length is the distance between the lens and the focal point. When the focal point is on the sensor, the image is in focus.



Focal length is related to angle of view in that a long focal length corresponds to a small angle of view, and vice versa.

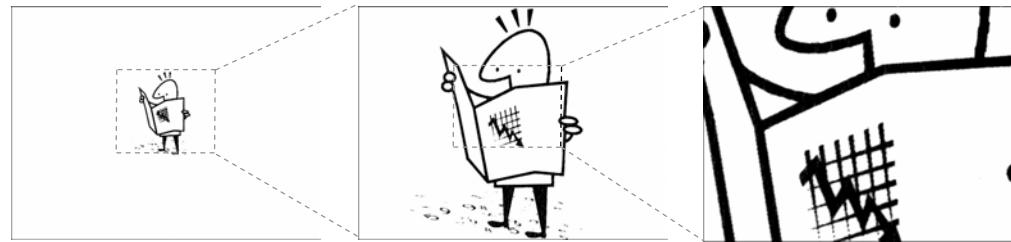
Example

Image taken with a wide angle lens, i.e. having a small focal length (8mm).

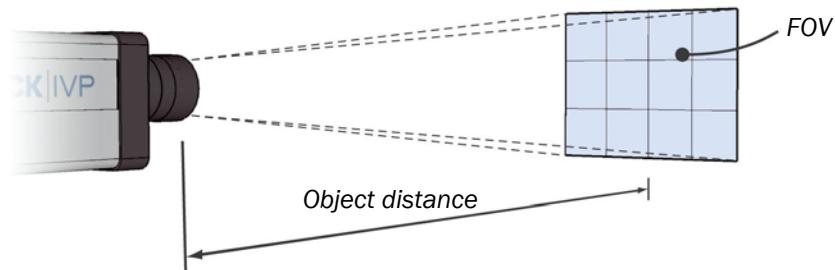
Image taken from the same distance with a medium focal length (25 mm)

Image taken from the same distance with a long focal length (50 mm tele).

In addition to the standard lenses there are other types for special purposes, described in further detail in the Appendix.

2.1.3 Field of View in 2D

The **FOV** (Field of View) in **2D** systems is the full area that a camera sees. The FOV is specified by its width and height. The **object distance** is the distance between the lens and the object.



The object distance is also called LTO (lens-to-object) distance or **working distance**.

2.1.4 Aperture and F-stop

The **aperture** is the opening in the lens that controls the amount of light that is let onto the sensor. In quality lenses, the aperture is adjustable.



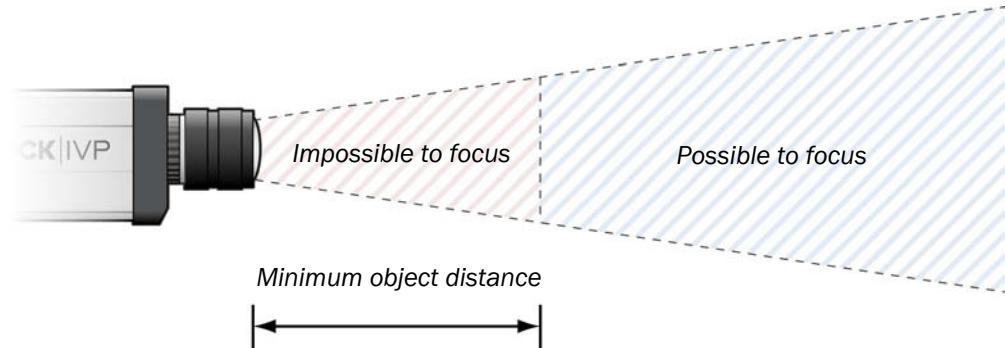
Large aperture, much light is let through.

Small aperture, only lets a small amount of light through.

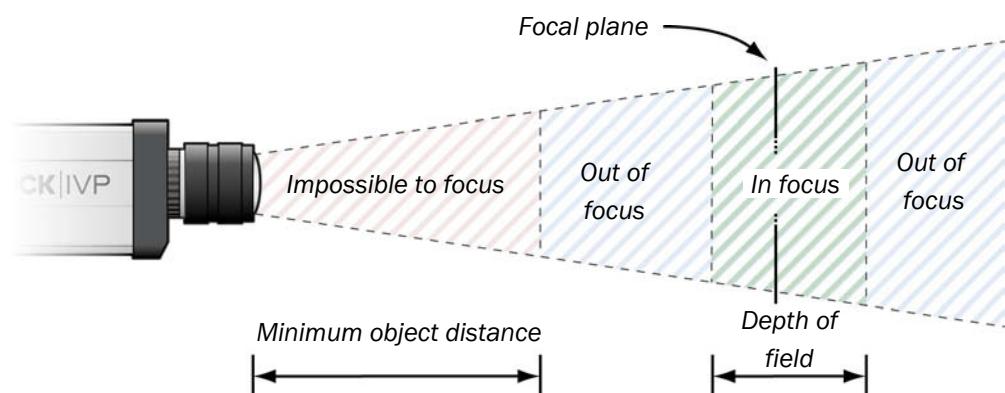
The size of the aperture is measured by its **F-stop** value. A large F-stop value means a small aperture opening, and vice versa. For standard CCTV lenses, the F-stop value is adjustable in the range between F1.4 and F16.

2.1.5 Depth of Field

The **minimum object distance** (sometimes abbreviated MOD) is the closest distance in which the camera lens can focus and **maximum object distance** is the farthest distance. Standard lenses have no maximum object distance (infinity), but special types such as macro lenses do.



The **focal plane** is found at the distance where the focus is as sharp as possible. Objects closer or farther away than the focal plane can also be considered to be in focus. This distance interval where good-enough focus is obtained is called **depth of field (DOF)**.



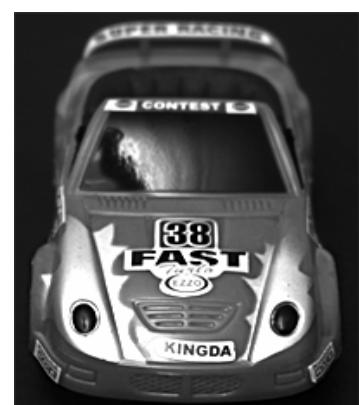
The depth of field depends on both the focal length and the aperture adjustment (described in next section). Theoretically, perfect focus is only obtained in the focal plane at an exact distance from the lens, but for practical purposes the focus is good enough within the depth of field. Rules of thumb:

1. A long focal length gives a shallow depth of field, and vice versa.
2. A large aperture gives a shallow depth of field, and vice versa.

Example



Small aperture and deep depth of field.



*Large aperture and shallow depth of field.
Notice how the far-away text is blurred.*

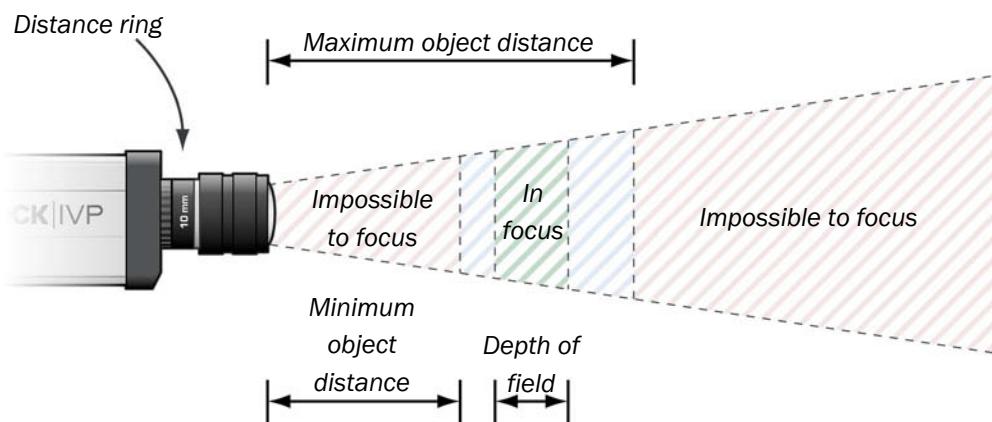
By adding a **distance ring** between the camera and the lens, the focal plane (and thus the MOD) can be moved closer to the camera. A distance ring is also referred to as **shim**, **spacer**, or **extension ring**.

A thick distance ring is called an **extension tube**. It makes it possible to position the camera very close to the object, also known as **macro** functionality.



Distance rings and extension tubes are used to decrease the minimum object distance. The thicker the ring or tube, the smaller the minimum object distance.

A side-effect of using a distance ring is that a **maximum object distance** is introduced and that the depth of field range decreases.



2.2 Basic Image Concepts

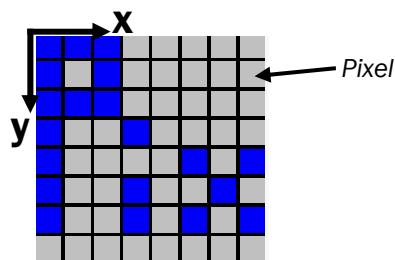
This section treats basic image terminology and concepts that are needed when working with any vision sensor or system.

2.2.1 Pixels and Resolution

A **pixel** is the smallest element in a digital image. Normally, the pixel in the image corresponds directly to the physical pixel on the sensor.

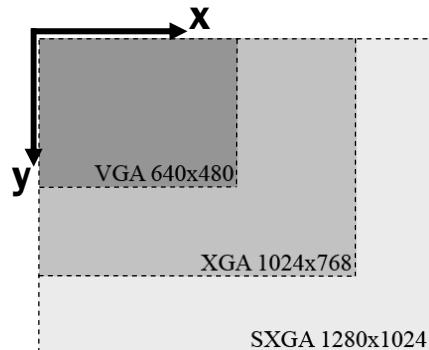
Pixel is an abbreviation of 'picture element'. Normally, the pixels are so small that they only become distinguishable from one another if the image is enlarged.

To the right is an example of a very small image with dimension 8x8 pixels. The dimensions are called **x** and **y**, where x corresponds to the image **columns** and y to the **rows**.



Typical values of **sensor resolution** in 2D machine vision are:

1. **VGA** (Video Graphics Array): 640x480 pixels
2. **XGA** (Extended Graphics Array): 1024x768 pixels
3. **SXGA** (Super Extended Graphics Array): 1280x1024 pixels



Note the direction of the y axis, which is opposite from what is taught in school mathematics. This is explained by the image being treated as a matrix, where the upper-left corner is the (0,0) element. The purpose of the coordinate system and matrix representation is to make calculations and programming easier.

The **object resolution** is the physical dimension on the object that corresponds to one pixel on the sensor. Common units for object resolution are μm (microns) per pixel and mm per pixel. In some measurements the resolution can be smaller than a pixel. This is achieved by interpolation algorithms that extract **subpixel** information from pixel data.

Example: Object Resolution Calculation

The following practical method gives a good approximation of the object resolution:

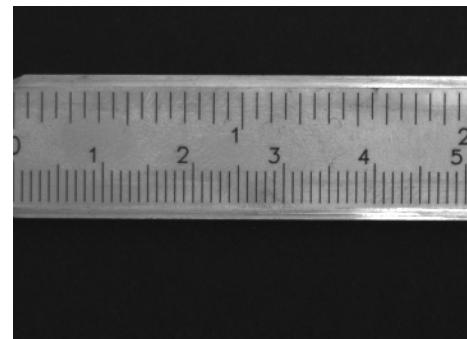
FOV width = 50 mm

Sensor resolution = 640x480 pixels

Calculation of object resolution in x:

$$= \frac{50}{640} = 0.08 \text{ mm / pix}$$

Result: The object resolution is 0.08 mm per pixel in x.



2.2.2 Intensity

The brightness of a pixel is called **intensity**. The intensity information is stored for each pixel in the image and can be of different types. Examples:

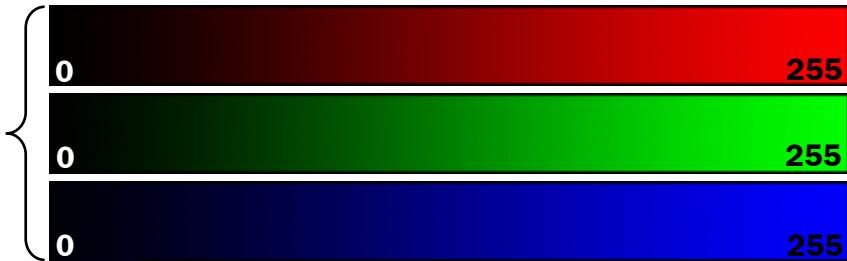
1. **Binary**: One bit per pixel.



2. **Gray scale**: Typically one byte per pixel.



3. **Color:** Typically one byte per pixel and color. Three bytes are needed to obtain full color information. One pixel thus contains three components (**R**, **G**, **B**).



When the intensity of a pixel is digitized and described by a byte, the information is quantized into discrete levels. The number of bits per byte is called **bit-depth**. Most often in machine vision, 8 bits per pixel are enough. Deeper bit-depths can be used in high-end sensors and sensitive applications.

Example



Binary image.



Gray scale image.



Color image.

Because of the different amounts of data needed to store each pixel (e.g. 1, 8, and 24 bits), the image processing time will be longer for color and gray scale than for binary.

2.2.3 Exposure

Exposure is how much light is detected by the photographic film or sensor. The exposure amount is determined by two factors:

1. **Exposure time:** Duration of the exposure, measured in milliseconds (ms). Also called **shutter time** from traditional photography.
2. **Aperture size:** Controls the amount of light that passes through the lens.

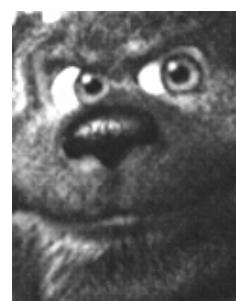
Thus the total exposure is the combined result of these two parameters.

If the exposure time is too short for the sensor to capture enough light, the image is said to be **underexposed**. If there is too much light and the sensor is **saturated**, the image is said to be **overexposed**.

Example



Underexposed image.



Normally exposed image.



Overexposed image with saturated areas (white).

A topic related to exposure time is **motion blur**. If the object is moving and exposure time is too long, the image will be blurred and application robustness is threatened. In applications where a short exposure time is necessary because of object speed, there are three methods to make the image bright enough:

1. Illuminate the object with high-intensity lighting (strobe)
2. Open up the aperture to allow more light into the camera
3. Electronic gain (described in next section)

Example



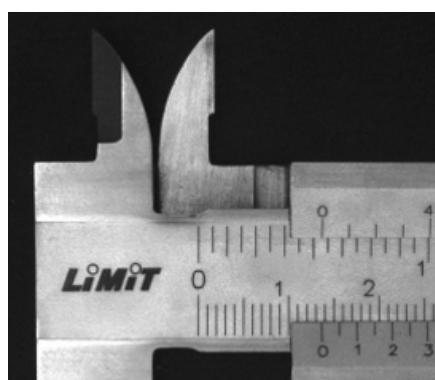
A short exposure time yields a sharp image.



A long exposure time causes motion blur if the object is moving fast.

2.2.4 Gain

Exposure time and aperture size are the physical ways to control image intensity. There is also an electronic way called **gain** that amplifies the intensity values after the sensor has already been exposed, very much like the volume control of a radio (which doesn't actually make the artist sing louder). The tradeoff of compensating insufficient exposure with a high gain is amplified **noise**. A grainy image appears.



Normally exposed image.

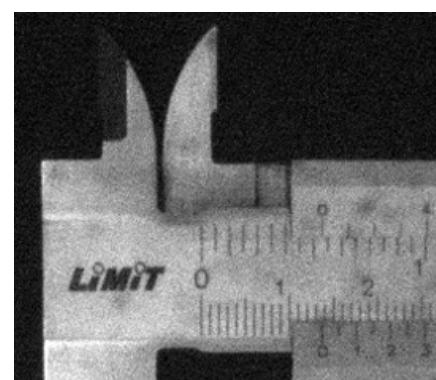


Image where underexposure has been compensated with a high gain.

2.2.5 Contrast and Histogram

Contrast is the relative difference between bright and dark areas in an image. Contrast is necessary to see anything at all in an image.



Low contrast.

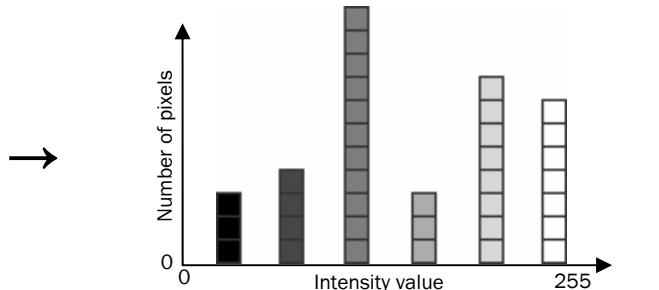
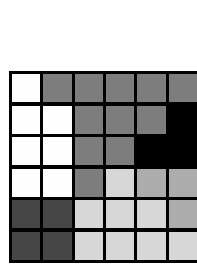


Normal contrast.



High contrast.

A **histogram** is a diagram where the pixels are sorted in order of increasing intensity values. Below is an example image that only contains six different gray values. All pixels of a specific intensity in the image (left) become one bar in the histogram (right).



Example image.

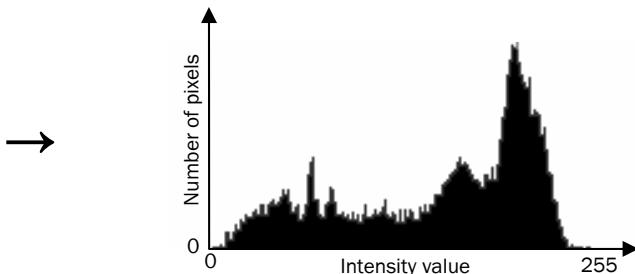
Histogram of the example image.

Histograms for color images work the same way as for grayscale, where each color channel (**R**, **G**, **B**) is represented by its individual histogram.

Typically the gray scale image contains many more gray levels than those present in the example image above. This gives the histogram a more continuous appearance. The histogram can now be used to understand the concept of contrast better, as shown in the example below. Notice how a lower contrast translates into a narrower histogram.



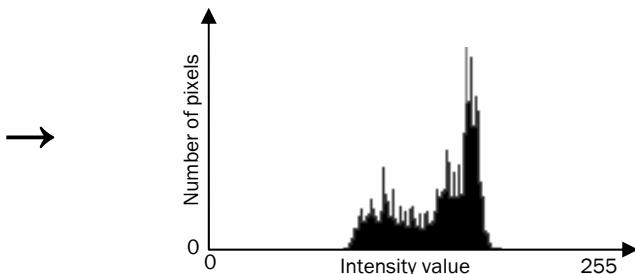
Normal contrast.



The histogram covers a large part of the gray scale.



Low contrast.



The histogram is compressed.

3**Illumination**

Light is of crucial importance in machine vision. The goal of **lighting** in machine vision is to obtain a robust application by:

1. Enhancing the features to be inspected.
2. Assuring high repeatability in image quality.

Illumination is the way an object is lit up and lighting is the actual lamp that generates the illumination.

Light can be **ambient**, such as normal indoor light or sunlight, or **special** light that has been chosen with the particular vision application's needs in mind.

Most machine vision applications are sensitive to lighting variations, why ambient light needs to be eliminated by a cover, called **shroud**.

3.1 Illumination Principles

Using different illumination methods on the same object can yield a wide variety of results. To enhance the particular features that need to be inspected, it is important to understand basic illumination principles.

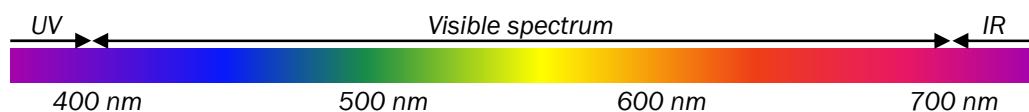
3.1.1 Light and Color

Light can be described as waves with three properties:

1. Wavelength or color, measured in nm (nanometers)
2. Intensity
3. Polarization.

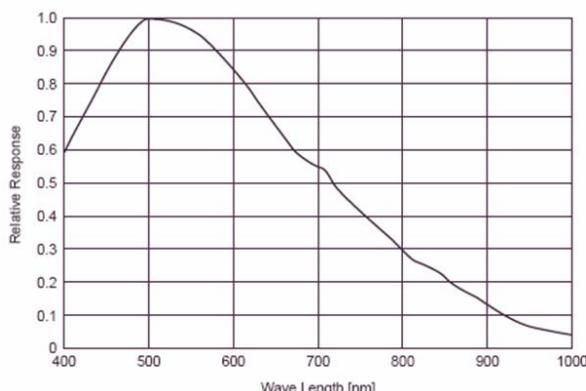
Mainly wavelength and intensity is of importance in machine vision, whereas the polarization is only considered in special cases.

Different wavelengths correspond to different colors. The human eye can see colors in the **visible spectrum**, whose colors range from violet to red. Light with shorter wavelength than violet is called **UV** (ultraviolet) and longer wavelength than red is called **IR** (infrared).



The **spectral response** of a sensor is the sensitivity curve for different wavelengths. Camera sensors can have a different spectral response than the human eye.

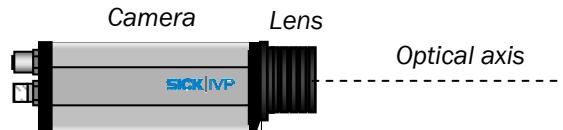
Example



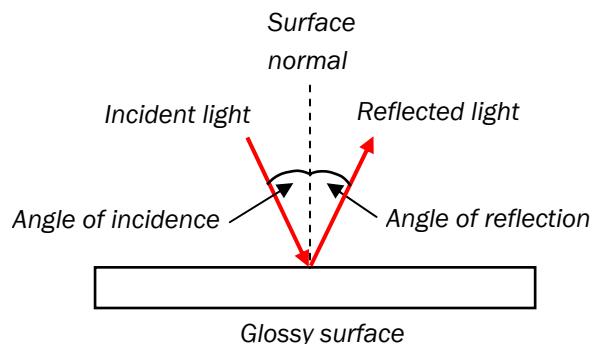
Spectral response of a gray scale CCD sensor. Maximum sensitivity is for green (500 nm).

3.1.2 Reflection, Absorption, and Transmission

The **optical axis** is a thought line through the center of the lens, i.e. the direction the camera is looking.

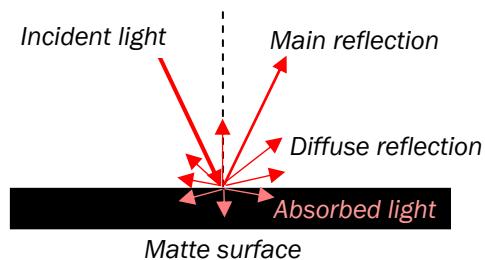


The camera sees the object thanks to light that is **reflected** on its surface. In the figure below, all light is reflected in one direction. This is called **direct** or **specular reflection** and is most prevalent when the object is glossy (mirror-like).

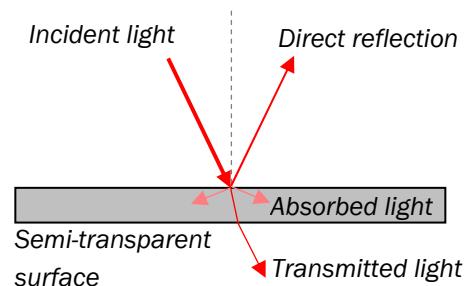


The **angle of incidence** and **angle of reflection** are always equal when measured from the surface normal.

When the surface is not glossy, i.e. has a **matte** finish, there is also **diffuse reflection**. Light that is not reflected is **absorbed** in the material.



Transparent or semi-transparent materials also **transmit** light.



The above principles, reflection, absorption, and transmission, constitute the basis of most lighting methods in machine vision.

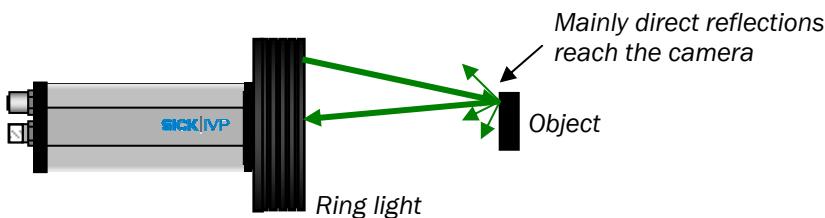
There is a fourth principle, **emission**, when a material produces light, for example when molten steel glows red because of its high temperature.

3.2 Lighting Types

There is a large variety of different lighting types that are available for machine vision. The types listed here represent some of the most commonly used techniques. The most accepted type for machine vision is the **LED** (Light-Emitting Diode), thanks to its even light, long life, and low power consumption.

3.2.1 Ring Light

A **ring light** is mounted around the optical axis of the lens, either on the camera or somewhere in between the camera and the object. The angle of incidence depends on the ring diameter, where the lighting is mounted, and at what angle the LEDs are aimed.



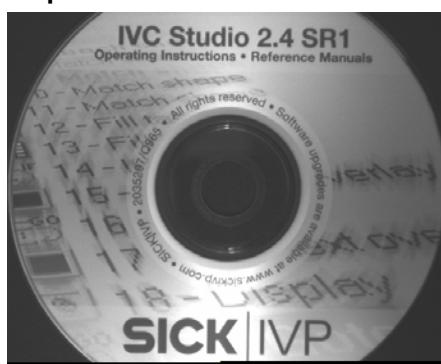
Pros

- Easy to use
- High intensity and short exposure times possible

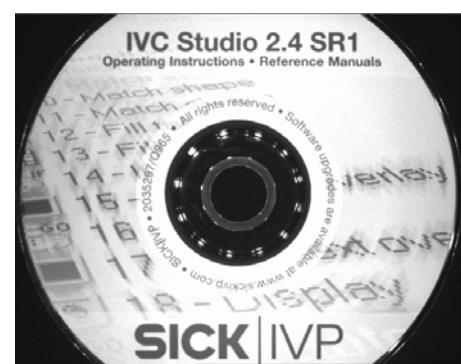
Cons

- Direct reflections, called **hot spots**, on reflective surfaces

Example



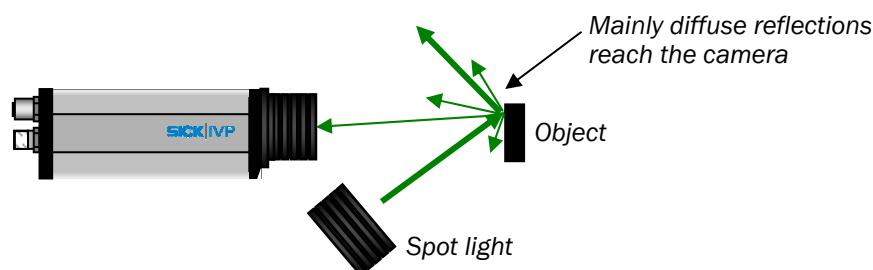
Ambient light.



Ring light. The printed matte surface is evenly illuminated. Hot spots appear on shiny surfaces (center), one for each of the 12 LEDs of the ring light.

3.2.2 Spot Light

A **spot light** has all the light emanating from one direction that is different from the optical axis. For flat objects, only diffuse reflections reach the camera.



Pros

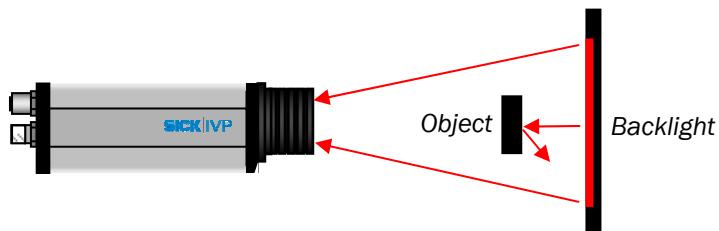
- No hot spots

Cons

- Uneven illumination
- Requires intense light since it is dependent on diffuse reflections

3.2.3 Backlight

The **backlight** principle has the object being illuminated from behind to produce a contour or silhouette. Typically, the backlight is mounted perpendicularly to the optical axis.

**Pros**

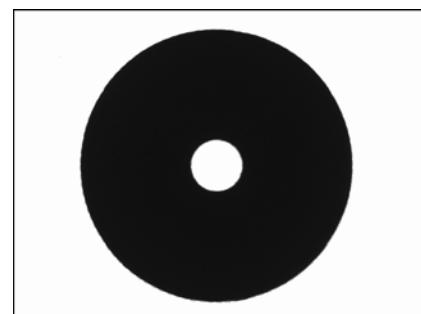
- Very good contrast
- Robust to texture, color, and ambient light

Cons

- Dimensions must be larger than object

Example

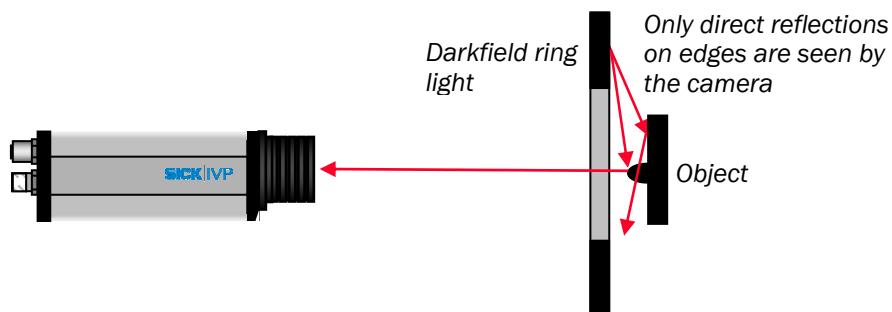
Ambient light.



Backlight: Enhances contours by creating a silhouette.

3.2.4 Darkfield

Darkfield means that the object is illuminated at a large angle of incidence. Direct reflections only occur where there are edges. Light that falls on flat surfaces is reflected away from the camera.



Pros

- Good enhancement of scratches, protruding edges, and dirt on surfaces

Cons

- Mainly works on flat surfaces with small features
- Requires small distance to object
- The object needs to be somewhat reflective

Example

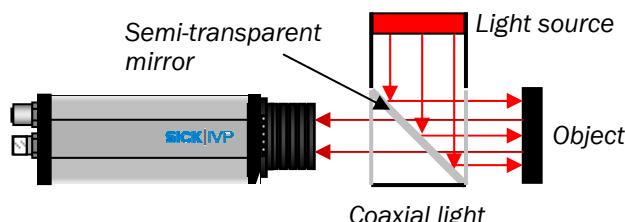
Ambient light.



Darkfield: Enhances relief contours, i.e. lights up edges.

3.2.5 On-Axis Light

When an object needs to be illuminated parallel to the optical axis, i.e. directly from the front, a semi-transparent mirror is used to create an **on-axis light** source. On-axis is also called **coaxial**. Since the beams are parallel to the optical axis, direct reflexes appear on all surfaces parallel to the focal plane.

**Pros**

- Very even illumination, no hot spots
- High contrast on materials with different reflectivity

Cons

- Low intensity requires long exposure times
- Cleaning of semi-transparent mirror (beam-splitter) often needed

Example

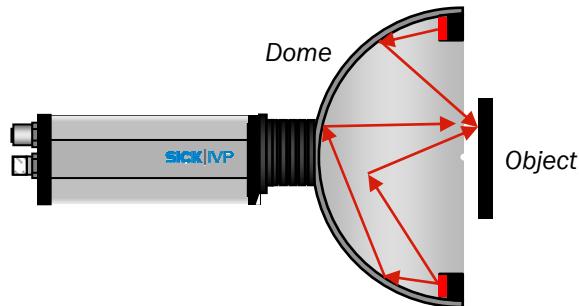
Inside of a can as seen with ambient light.



Inside of the same can as seen with a coaxial (on-axis) light:

3.2.6 Dome Light

Glossy materials can require a very diffuse illumination without hot spots or shadows. The **dome light** produces the needed uniform light intensity thanks to LEDs illuminating the bright, matte inside of the dome walls. The middle of the image becomes darker because of the hole in the dome through which the camera is looking.



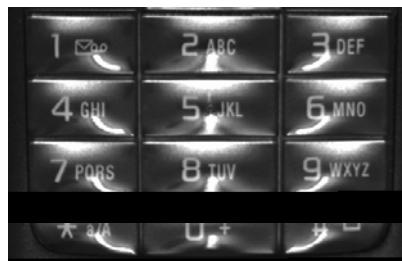
Pros

- Works well on highly reflective materials
- Uniform illumination, except for the darker middle of the image. No hot spots

Cons

- Low intensity requires long exposure times
- Dimensions must be larger than object
- Dark area in the middle of the image

Example



Ambient light. On top of the key numbers is a curved, transparent material causing direct reflections.



The direct reflections are eliminated by the dome light's even illumination.

3.2.7 Laser Line

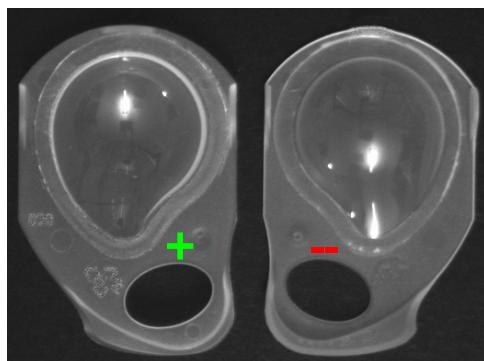
Low-contrast and 3D inspections normally require a 3D camera. In simpler cases where accuracy and speed are not critical, a 2D camera with a laser line can provide a cost-efficient solution.

Pros

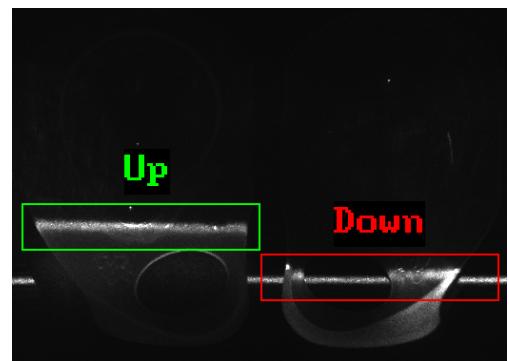
- Robust against ambient light.
- Allows height measurements (z parallel to the optical axis).
- Low-cost 3D for simpler applications.

Cons

- Laser safety issues.
- Data along y is lost in favor of z (height) data.
- Lower accuracy than 3D cameras.

Example

Ambient light. Contact lens containers, the left is facing up (5 mm high at cross) and the right is facing down (1 mm high at minus sign).



The laser line clearly shows the height difference.

3.3 Lighting Variants and Accessories

3.3.1 Strobe or Constant Light

A **strobe light** is a flashing light. Strobing allows the LED to emit higher light intensity than what is achieved with a **constant light** by **turbo charging**. This means that the LED is powered with a high current during on-time, after which it is allowed to cool off during the off-time. The on-time relative to the total cycle time (on-time plus off-time) is referred to as **duty cycle (%)**.

With higher intensity, the exposure time can be shortened and motion blur reduced. Also, the **life** of the lamp is extended.

Strobing a LED lighting requires both software and hardware support.

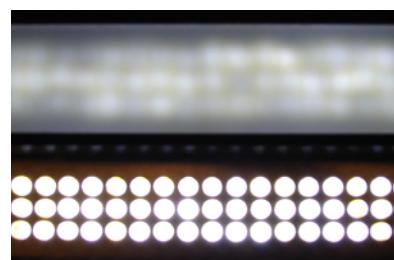
3.3.2 Diffusor Plate

Many lighting types come in two versions, with or without a **diffusor plate**. The diffusor plate converts direct light into diffuse.

The purpose of a diffusor plate is to avoid bright spots in the image, caused by the direct light's reflections in glossy surfaces.

Rules of thumb:

1. Glossy objects require diffuse light.
2. Diffusor plates steal light intensity.



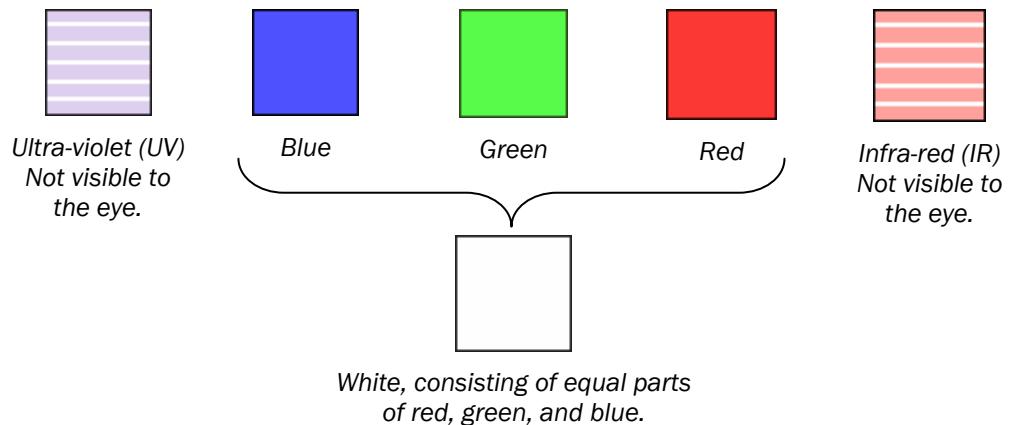
Two identical white bar lights, with diffusor plate (top) and without (bottom).

Typically, 20-40% of the intensity is lost in the diffusor plate, which can be an issue in high-speed applications where short exposure times are needed.

Diffusor plates work well on multi-LED arrays, whereas single LEDs will still give bright "hot spots" in the image.

3.3.3 LED Color

LED lightings come in several **colors**. Most common are red and green. There are also LEDs in blue, white, UV, and IR. Red LEDs are the cheapest and have up to 10 times longer life than blue and green LEDs.



Different objects reflect different colors. A blue object appears blue because it reflects the color blue. Therefore, if blue light is used to illuminate a blue object, it will appear bright in a gray scale image. If a red light is used to illuminate a blue object it will appear dark. It is thus possible to use color to an advantage, even in gray scale imaging.

3.3.4 Optical Filters

An **optical filter** is a layer in front of the sensor or lens that absorbs certain wavelengths (colors) or polarizations. For example, sunglasses have an optical filter to protect your eyes from hazardous UV radiation. Similarly, we can use a filter in front of the camera to keep the light we want to see and suppress the rest.



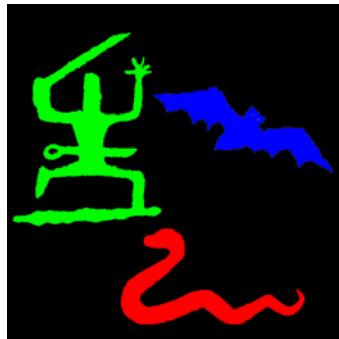
Two main optical filter types are used for machine vision:

1. **Band-pass filter:** Only transmits light of a certain color, i.e. within a certain wavelength interval. For example, a red filter only lets red through.
2. **Polarization filter:** Only transmits light with a certain polarization. Light changes its polarization when it is reflected, which allows us to filter out unwanted reflections.

Very robust lighting conditions can be achieved by combining an appropriate choice of LED color with an optical filter having the corresponding band-pass wavelength.

Example

By combining optical filters and selected LED colors, it is possible to improve contrast between an object of a certain color in the image and its surrounding.



Original image.



Image seen by gray scale camera with ambient light and without filter.



Red light and a red band-pass filter.



Green light and a green band-pass filter.

3.4 Safety and Eye Protection

Light can be harmful if the intensity is too high. Lightings for vision sometimes reach harmful intensity levels, especially in techniques where lasers are used but sometimes also for LEDs. It is important to know the **safety classification** of the lighting before using it in practice.



Damage of the eye can be temporary or permanent, depending on the exposure amount. When the damage is permanent, the light-sensitive cells on the eye's retina have died and will not grow back. The resulting **blindness** can be partial or total, depending on how much of the retina that has been damaged.

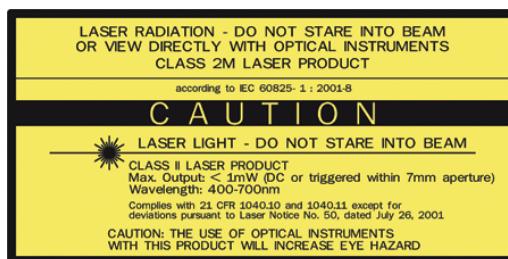
3.4.1 Laser Safety

A laser is a light source that emits parallel light beams of one wavelength (color), which makes the laser light dangerous for the eye.

Lasers are classified into **laser classes**, ranging from 1 to 4. Classes 2 and 3 are most common in machine vision. Below is an overview of the classifications.

European class	American class	Practical meaning
1-1M	I	Harmless. Lasers of class 1M may become hazardous with the use of optics (magnifying glass, telescope, etc).
2-2M	II	Caution. Not harmful to the eye under normal circumstances. Blink reflex is fast enough to protect the eye from permanent damage. Lasers of class 2M may become hazardous with the use of optics.
3R-3B	IIIb	Danger. Harmful at direct exposure of the retina or after reflection on a glossy surface. Usually doesn't produce harmful diffuse reflections.
4	IV	Extreme danger. With hazardous diffuse reflections.

Example



Example of warning label for laser class II/2M.

3.4.2 LEDs

LEDs are not lasers from a technical point of view, but they behave similarly in that they have a small size and emit light in one main direction. Because of this the intensity **can be harmful** and a safety classification is needed. There is no system for classifying LEDs specifically, so the laser classification system has been (temporarily?) adopted for LEDs.

LEDs are often used in **strobe lights**, which can cause **epileptic seizures** at certain frequencies in people with epilepsy.

3.4.3 Protective Eyewear

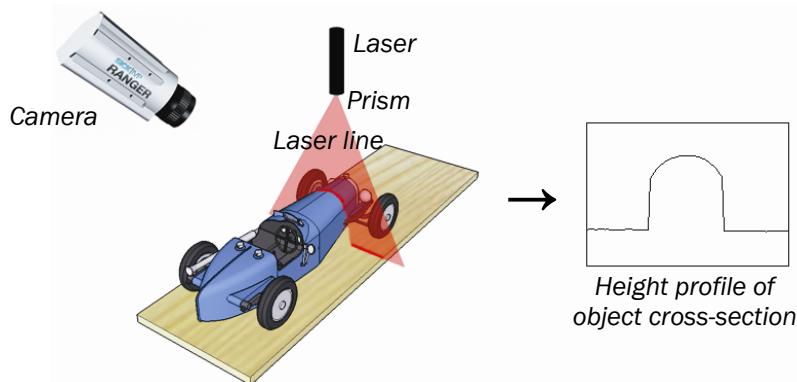
Protective eyewear is necessary whenever working with dangerous light. Their purpose is to absorb enough light so that the intensity becomes harmless to the eye. Three aspects are important when choosing safety goggles:

1. Which wavelength is emitted by the laser/LED?
2. Which wavelengths are absorbed by the goggles?
3. How much of the light intensity is absorbed?



4**Laser Triangulation**

Laser triangulation is a technique for acquiring 3D height data by illuminating the object with a **laser** from one direction and having the camera look from another. The laser beam is divided into a **laser line** by a **prism**. The **view angle** makes the camera see a **height profile** that corresponds to the object's **cross-section**.



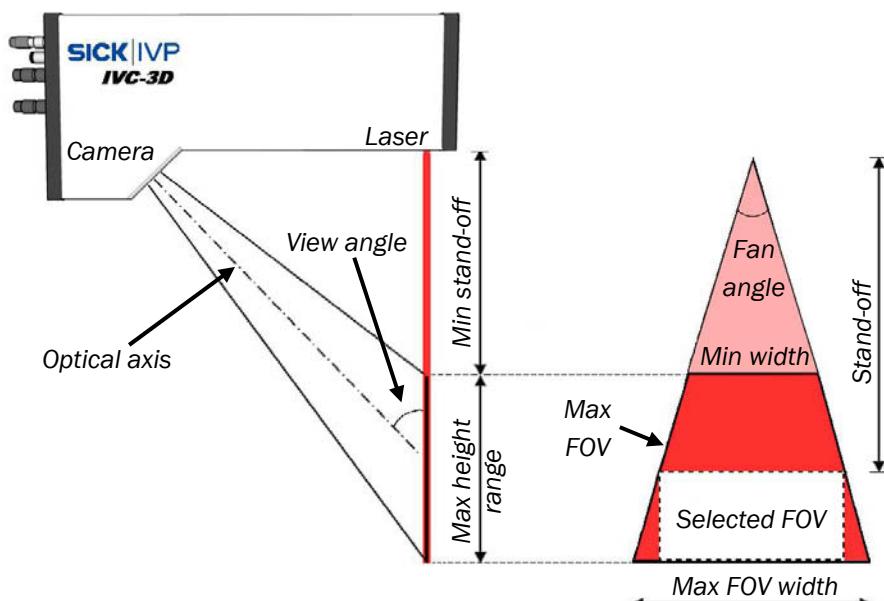
A laser line is projected onto the object so that a height profile can be seen by a camera from the side. The height profile corresponds to the cross-section of the object.

The method to grab a complete **3D image** is to move the object under the laser line and put together many consecutive profiles.

4.1 Field of View in 3D

The **selected FOV** (Field of View) is the rectangular area in which the camera sees the object's cross-section. The selected FOV, also called defining rectangle, lies within the trapezoid-shaped maximum FOV.

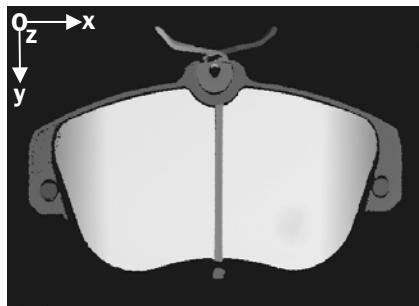
There are several possible camera/laser geometries in laser triangulation. In the basic geometry, the distance between the camera unit and the top of the FOV is called **stand-off**. The possible width of the FOV is determined by the focal length of the lens, the laser prism's **fan angle**, and the stand-off.



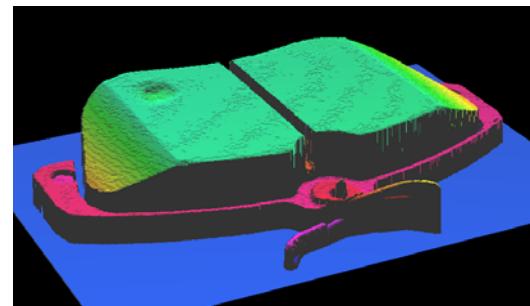
The fan angle of the laser line gives the maximum FOV a trapezoidal shape. Within this, the selected FOV defines the cross-section where the camera is looking at the moment.

4.2 3D Image and Coordinate System

There are a number of different representations of 3D data. SICK IVP uses intensity-coded height data, where bright is high and dark is low. This can be transformed into a 3D visualization with color-coded height data.



3D image with intensity-coded height data.



3D image visualized in 3D viewer with color-coded height data.

The coordinate system in the 3D image is the same as that of a normal 2D image regarding x and y, with the addition that y now corresponds to time. The additional height dimension is referred to as the **z axis** or the **range axis**.

Since the front of the object becomes the first scanned row in the image, the y axis will be directed opposite to the conveyor movement direction.

4.3 Scanning Speed

Since laser triangulation is a line scanning method, where the image is grabbed little by little, it is important that the object moves in a controlled way during the scan.

This can be achieved by either:

1. An **encoder** that gives a signal each time the conveyor has moved a certain distance.
2. A **constant conveyor speed**.

When an encoder is used, it controls the **profile triggering** so that the profiles become equidistant.

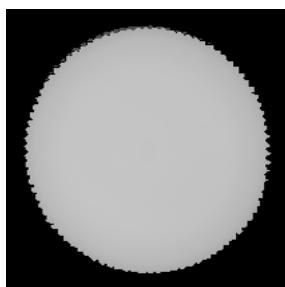
A constant conveyor speed can often not be guaranteed, why an encoder is generally recommended.



Encoder.

It is important to note that there is a maximum speed at which the profile grabbing can be done, determined by the maximum **profile rate** (profiles/second).

If this speed or maximum profile rate is exceeded, some profiles will be lost and the image will be **distorted** despite the use of an encoder. A distorted image means that the object proportions are wrong. An image can also appear distorted if the x and y resolution are different (i.e. non-square pixels), which can be desirable when optimizing the resolution.



3D image of a circular object. The proportions are correct thanks to the use of an encoder.



A distorted 3D image of the same object. The proportions are incorrect despite the use of an encoder, because the scanning speed has exceeded the maximum allowed profile rate.

The maximum profile rate is limited by three main factors:

1. The **exposure** of the sensor. A longer exposure time per profile reduces the maximum profile rate.
2. The sensor **read-out time**. The time it takes to convert the sensor information to a digital format.
3. The **data transfer rate**. The time it takes to transfer 3D data from the camera to the signal processing unit.

To ensure that the object is scanned in its entirety, it is common to use a **photo switch** to start the acquisition at the correct moment. The photo switch is thus used for **image triggering**.

In some applications where there is a more continuous flow on the conveyor, it is not meaningful to trig the scanning by a photo switch. Instead, the camera is used in **free-running** mode, which means that the acquisition of a new image starts as soon as the previous image is completed. Sometimes it is necessary to have **overlapping images** to ensure that everything on the conveyor is fully scanned and analyzed.

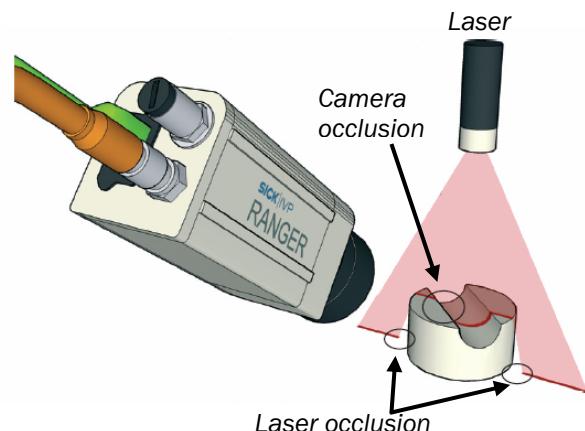
4.4 Occlusion and Missing Data

Because of the view angle between the camera and the laser line, the camera will not be able to see behind object features. This phenomenon is called **camera occlusion** (shadowing) and results in **missing data** in the image.

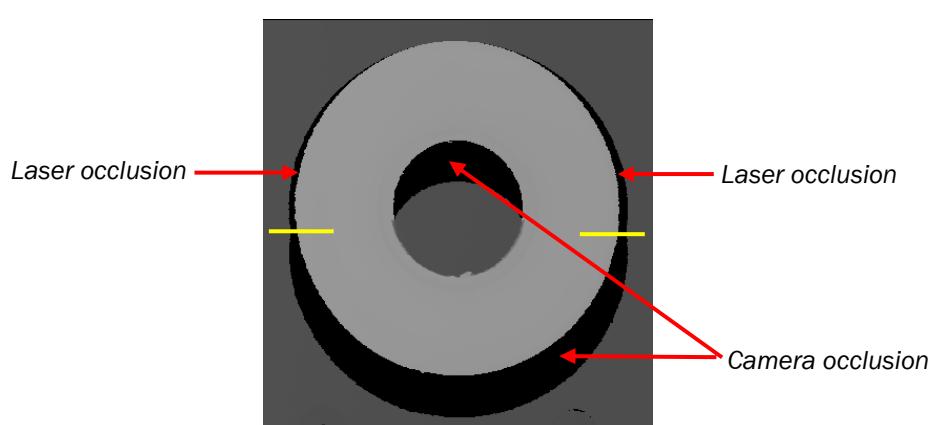
As a consequence, laser triangulation is not a suitable for scanning parts of an object located behind high features. Examples are inspections of the bottom of a hole or behind steep edges.

Because of the fan angle of the laser, the laser line itself can be occluded and result in missing data. This phenomenon is called **laser occlusion**.

The image below of a roll of scotch tape shows both camera and laser occlusion. The yellow lines show where the camera occlusion starts to dominate over the laser occlusion.



*Camera occlusion occurs behind features as seen from the camera's perspective.
Laser occlusion occurs behind features as seen from the laser's perspective.*

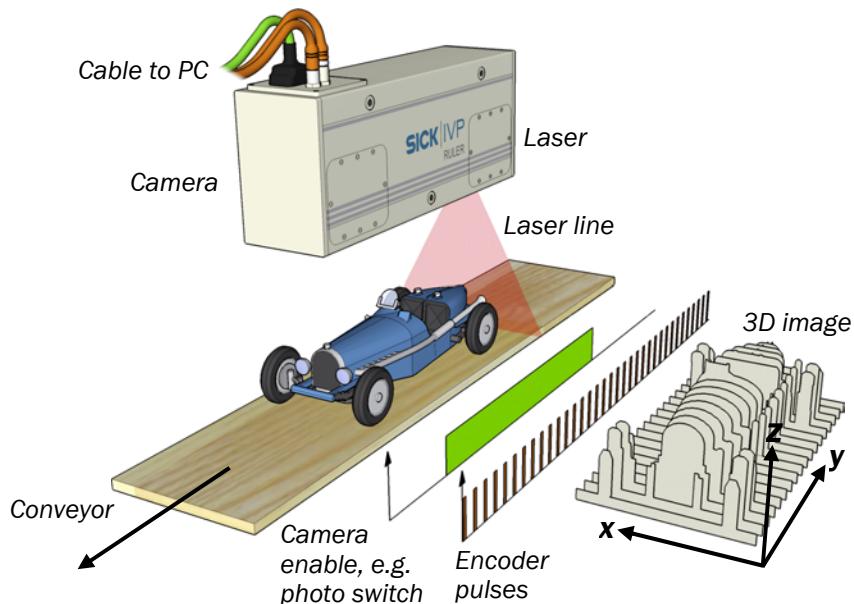


Intensity-coded 3D image of a roll of scotch tape, showing both camera and laser occlusion.

4.5 System Components

A typical laser triangulation setup consists of the following components:

1. A **laser** to produce the height profile.
2. A **camera** to scan profiles.
3. A **conveyor** to move the object under the camera.
4. A **photo switch** to enable the camera when an object is present.
5. An **encoder** to ensure that the profiles are grabbed at a constant distance, independent of conveyor speed (up to its maximum allowed value).
6. An **image processing unit**, either built-in (smart camera) or external (PC), to collect profiles into an image and to analyze the result.



In some laser triangulation products, all of the above components are bought and configured separately for maximum flexibility. Others are partially assembled, for example with a fixed geometry (view angle), which makes them more ready to use but less flexible.

Example

1. **SICK IVP Ranger**: All components are separated.
2. **SICK IVP Ruler**: Camera and laser are built in to create a fixed geometry.
3. **SICK IVP IVC-3D**: Camera and laser are built in to create a fixed geometry. In addition to this, the unit contains both image processing hardware and software for stand-alone use.

4.6 Ambient Light Robustness

The laser emits monochromatic light, meaning that it only contains one wavelength. By using a narrow band-pass filter in front of the sensor, other wavelengths in the ambient light can be suppressed. The result is a system that is rather robust against ambient light. However, when the ambient light contains wavelengths close to that of the laser, this will be let through the filter and appear as disturbing **reflections** in the image. Then the installation needs to be covered, or **shrouded**. Typically, problems with reflections occur with **sunlight** and “warm” artificial light from **spotlights**.

5**Processing and Analysis**

After the image has been grabbed, the next step is image **analysis**. This is where the desired **features** are **extracted** automatically by **algorithms** and conclusions are drawn. A feature is the general term for information in an image, for example a dimension or a pattern. Algorithms are also referred to as **tools** or functions.

Sometimes the image needs **preprocessing** before the feature extraction, for example by using a digital **filter** to enhance the image.

5.1 Region of Interest

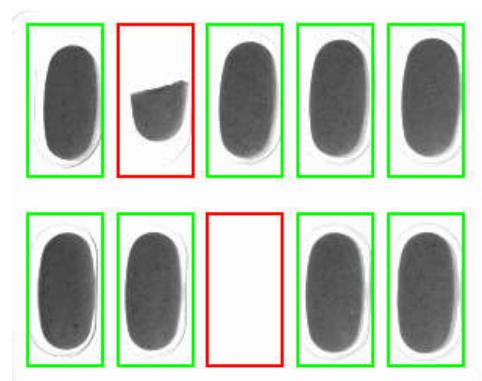
A **ROI** (Region of Interest) is a selected area of concern within an image. The purpose of using ROIs is to restrict the area of analysis and to allow for different analyses in different areas of the image. An image can contain any number of ROIs. Another term for ROI is **AOI** (Area of Interest).

A common situation is when the object location is not the same from image to image. In order to still inspect the feature of interest, a **dynamic ROI** that moves with the object can be created. The dynamic ROI can also be resized using results from previous analysis.

Examples



One ROI is created to verify the logotype (blue) and another is created for barcode reading (green).



A ROI is placed around each pill in the blister pack and the pass/fail analysis is performed once per ROI.

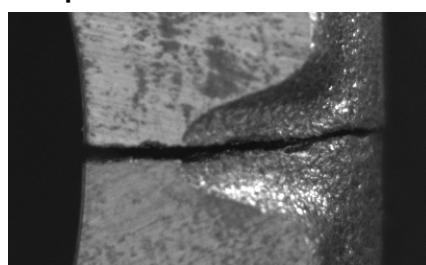
5.2 Pixel Counting

Pixel counting is the most basic analysis method. The algorithm finds the number of pixels within a ROI that have intensities within a certain gray level interval.

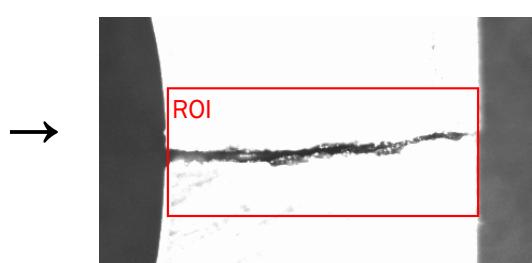
Pixel counting is used to measure **area** and to find **deviances** from a normal appearance of an object, for example missing pieces, spots, or cracks.

A pixel counter gives the **pixel sum** or area as a result.

Example



Automotive part with crack.



The crack is found using a darkfield illumination and by counting the dark pixels inside the ROI.

5.3 Digital Filters and Operators

Digital filtering and **operators** are used for preprocessing the image before the analysis to remove or enhance features. Examples are removal of noise and edge enhancement.

Examples



Original intensity-coded 3D image.

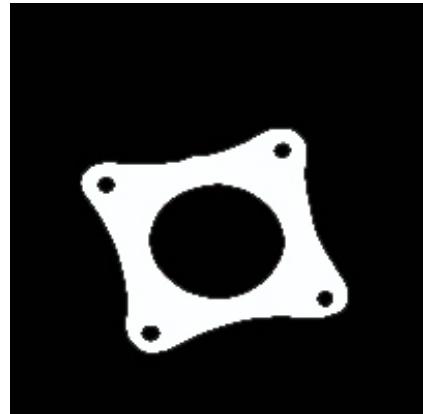
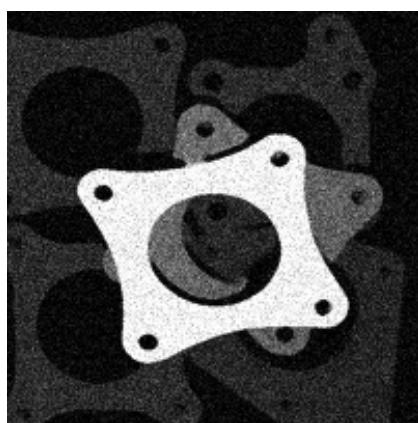


Image after a binarization operation.



Noisy version of original image.

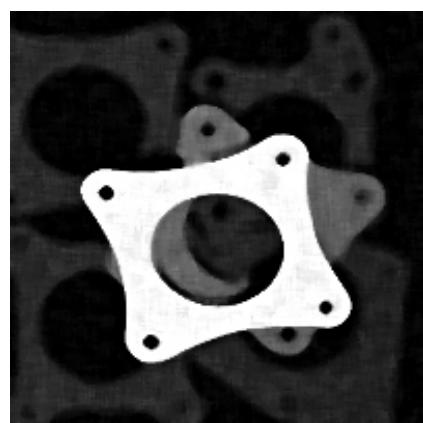
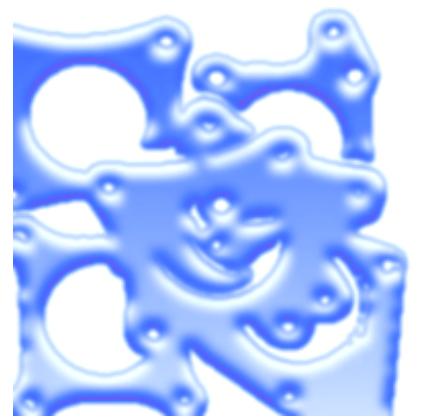


Image (left) after noise reduction.



Image after edge enhancement.



Example of artistic filtering,
with little or no use in machine vision.

5.4 Thresholds

A **threshold** is a limit. Thresholds can either be absolute or relative. In the context of gray scale images, an **absolute threshold** refers to a gray value (e.g. 0-255) and a **relative threshold** to a gray value difference, i.e. one gray value minus another.

A frequent use of thresholds is in **binarization** of gray scale images, where one absolute threshold divides the histogram into two intervals, below and above the threshold. All pixels below the threshold are made black and all pixels above the threshold are made white.

Absolute thresholds often appear in pairs as a **gray low** and a **gray high** threshold, to define closed gray scale intervals.

Example: Binarization



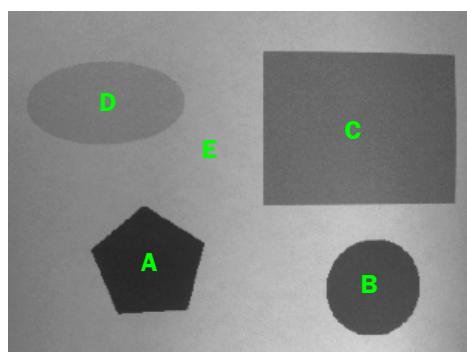
Example image: Gray scale.



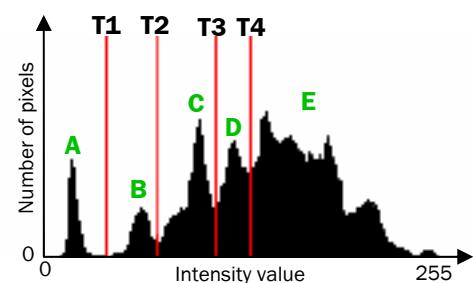
Binarized image: Binary.

Example: Double Absolute Thresholds

Objects **A** to **D** in the example image below can be separated from each other and from the background **E** by selecting a gray scale interval in the histogram. Each interval is defined by a gray low and a gray high threshold. Suitable thresholds **T1** to **T4** for separating the objects are drawn as red lines in the histogram.

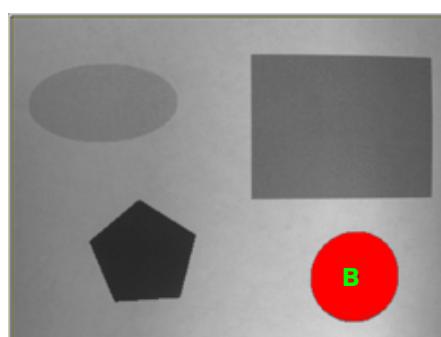


Example image.



Histogram of the example image.

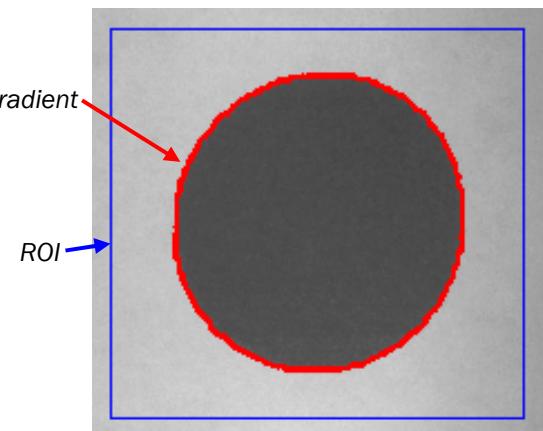
In the image below, object **B** is found by selecting gray low to **T1** and gray high to **T2**. The red coloring of the object highlights which pixels fall within the selected gray scale interval.



Example: Relative Threshold

Absolute thresholds are useful for finding areas of a certain gray scale whereas relative thresholds are useful to find **transitions**, or **edges**, where there is a gray scale **gradient** (change).

The image to the right shows the pixels where there is a gradient larger than a minimum relative threshold. If the threshold would have been too low, the algorithm would have found gradients on the noise level as well.

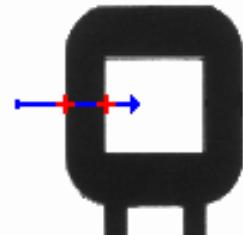


5.5 Edge Finding

An **edge** is defined by a change in intensity (2D) or height (3D). An edge is also called a **transition**. The task of an edge finding function is to extract the coordinates where the edge occurs, for example along a line.

Edge finding is used to **locate objects**, **find features**, and to **measure dimensions**.

An edge finder gives the **X** and **Y** coordinates as a result:



Edges (red crosses) are found along the search line.

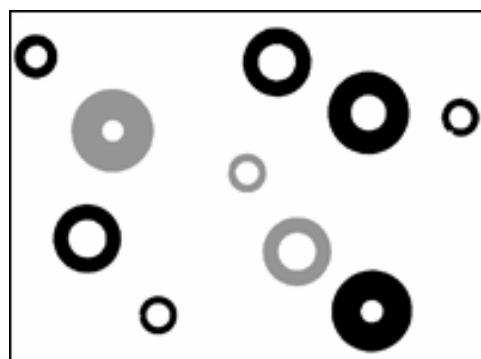
5.6 Blob Analysis

A **blob** is any area of **connected pixels** that fulfill one or more criteria, for example having a minimum area and intensity within a gray value interval.

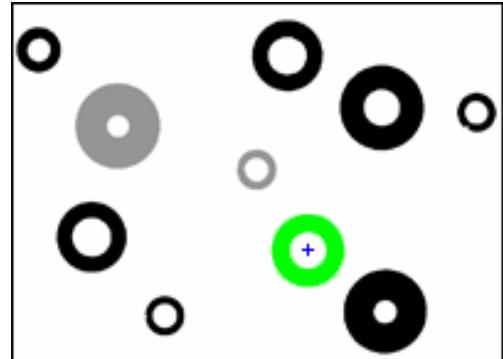
A **blob analysis** algorithm is used to find and count objects, and to make basic measurements of their characteristics.

Blob analysis tools can yield a variety of results, for example:

1. **Center of gravity**: Centroid. (Blue cross in the example)
2. **Pixel count**: Area. (Green pixels in the example)
3. **Perimeter**: Length of the line that encloses the blob area.
4. **Orientation**: Rotation angle.

Example

Example image: Blobs of four different sizes and two gray levels.



*Blob found by double search criteria:
Gray scale and area thresholding.*

5.7 Pattern Matching

Pattern matching is the recognition of a previously taught pattern in an image. Pattern matching can only be used when there is a **reference object** and the objects to inspect are (supposed to be) identical to the reference.

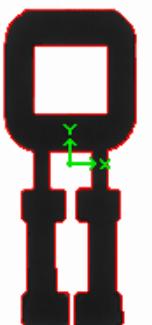
Pattern matching is used to locate objects, verify their shapes, and to align other inspection tools. The location of an object is defined with respect to a **reference point** (pickpoint) that has a constant position relative to the reference object.

Pattern matching algorithms for 2D can be either **gray scale** based (absolute) or **gradient** based (relative), which corresponds to height or height gradient based in 3D.

Pattern matching tools typically give the following results:

1. **X** and **Y** of reference point (pickpoint), and **Z** in 3D
2. **Orientation** (rotation)
3. **Match score** in % (likeness as compared to taught reference object)
4. **Number of found objects.**

Example



Reference image for teaching.
(Gradient-based algorithm.)

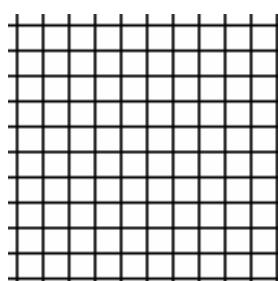


Matching in new image.

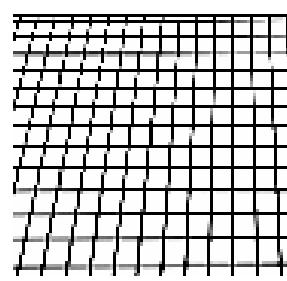
5.8 Coordinate Transformation and Calibration

Coordinate transformation converts between different coordinate systems, for example from image coordinates (x and y in pixels) to external, real-world coordinates (x, y, and z in mm for a robot). This procedure is also referred to as **calibration**.

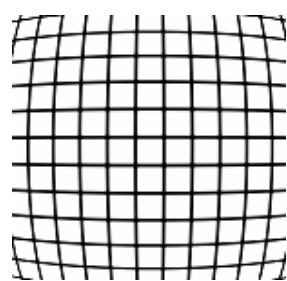
Coordinate transformation can be used to compensate for object rotation and translation, **perspective** (camera tilt), and lens **distortion**.



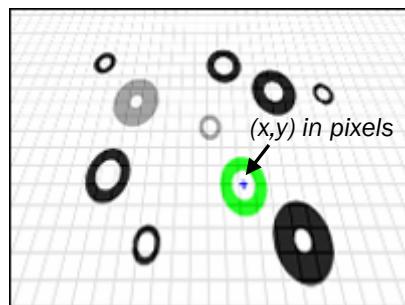
Real-world coordinates.



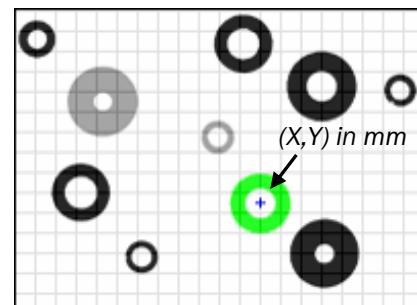
Perspective: Image seen by
a tilted camera.



Distortion: Image seen
through a wide-angle lens.

Example

Perspective in image.



Transformed image.

5.9 Code Reading

Codes are printed on products and packages to enable fast and automatic identification of the contents. The most common types are barcodes and matrix codes.

5.9.1 Barcode

A **barcode** is a 1D-code that contains **numbers** and consists of black and white vertical line elements. Barcodes are used extensively in packaging and logistics.

Examples of common barcode types are:

1. EAN-8 and EAN-13
2. Code 39 and Code 128
3. UPC-A and UPC-E
4. Interleaved 2 of 5.



121434656

Example of Code 39 barcode.

5.9.2 Matrix Code

A **matrix code** is a 2D-array of code elements (squares or dots) that can contain information of both **text and numbers**. The size of the matrix depends on how much information the code contains. Matrix codes are also known as **2D codes**.



Example of DataMatrix code.

An important feature with matrix codes is the **redundancy of information**, which means that the code is still fully readable thanks to a **correction scheme** although parts of the image are destroyed.

Examples of matrix codes are:

1. DataMatrix (e.g. with correction scheme ECC200)
2. PDF417
3. MaxiCode.

5.10 Text Verification and Reading

Automated text reading is used in packaging and logistics to inspect print quality and verify or read a printed message.

5.10.1 Optical Character Verification: OCV

Optical character verification, or **OCV**, is an algorithm that **verifies a taught-in text string**. The OCV function gives the result **true** (the correct string was found) or **false** (not found).

Example: OCV

OCV: Teach string.



OCV: Recognize string. A misprinted or incomplete character is identified as a fail.

5.10.2 Optical Character Recognition: OCR

Optical character recognition, or **OCR**, is an algorithm that reads or **recognizes unknown text**, where each letter is compared with a taught-in font. The OCR function gives the results:

1. The read **string**, i.e. the sequence of characters.
2. **True** or **false**, i.e. if the reading was successful or if one or more characters were not recognized as part of the font.

Two types of readers exist. One is the **fixed font reader** that uses fonts that are specially designed for use with readers. The other is the **flexible font reader** that in principle can learn any set of alphanumeric characters. For robustness of the application, however, it is important to choose a font where the characters are as different as possible from one another. Examples of a suitable font and a difficult one are:

1. **OCR A Extended**: In this font, similar characters have been made as dissimilar as possible, for example l and I, and the characters are equally spaced.
2. **Arial**: In this font, the similarity of certain characters can make it difficult or impossible for the algorithm, for example to distinguish between l and I (lower-case L and upper-case i). Tight distances between characters can also pose difficulties.

Example: OCR

OCR: Teach font.



OCR: Read string.

5.11 Cycle Time

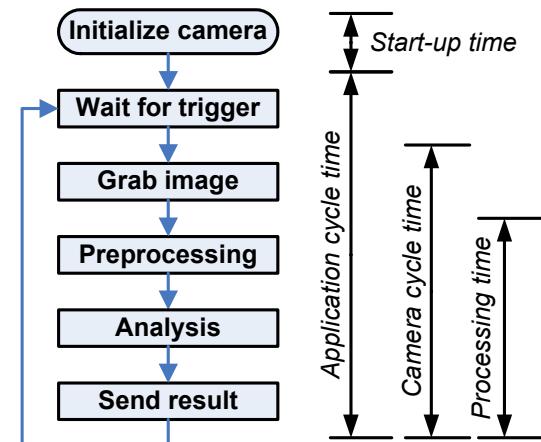
Vision systems that operate in automated production lines often need to be fast. The speed aspect of a vision system's performance is defined by its **cycle time**. The concept can be divided into subcategories, as illustrated by the flow diagram below.

The **start-up time** (or boot-up time) is the time from power-on to the point where the camera is ready to grab and analyze the first image.

The **application cycle time** of the application is the time between two consecutive inspections. It is equally common to state this in terms of **object frequency**, calculated as $1/(\text{application cycle time})$, which is the number of objects that pass the camera per second.

When the system runs at its maximum speed, the application cycle time will be the same as the minimum **camera cycle time**.

If the system runs faster than the camera cycle time can cope with, some objects will pass the inspection station uninspected.



The related term **processing time** (or **execution time**) refers to the time from the start of the analysis to the moment when the conclusion is drawn and the result is sent.

There are methods to optimize the cycle time with parallel grabbing and processing, which, in the best case, reduces the minimum cycle time to become equal to the processing time. This method is called **double buffering** or **ping-pong grabbing**.

A vision system's processes are usually timed in milliseconds (ms). The processing times for most applications are in the order of 10–500 ms.

5.12 Camera Programming

So far in this chapter, algorithms and methods have been described in what they do individually. Most applications are more complex in the sense that algorithms need to be combined and that one algorithm uses the result of another for its calculations. Achieving this requires a **programming environment**. This can be a specific, ready-to-use software for the camera, such as IVC Studio, or it can be a generic environment, such as Microsoft Visual Studio for more low-level C++ or Visual Basic programming.

A program can **branch** to do different things depending on intermediate or final results. This is obtained through **conditional instructions**, for example the **If** statement that is used for **pass/fail treatment**.

Calculations and string (text) operations are handled in the program by **expression evaluations**. An expression is a written formula that can contain numbers, variable references, or strings.

Depending on the type of the expression, the evaluation can give different results:

1. **Boolean**: 1 or 0, true or false, good or bad.
2. **Numeric**: a number, for example 3.1415.
3. **String**: text, for example "Best before May 2010" or "Lot code AAA".

A common situation is when a part of the program needs to be reused frequently. Instead of copying the code over and over again, it can be packaged in a **macro** that can be exported to other applications.

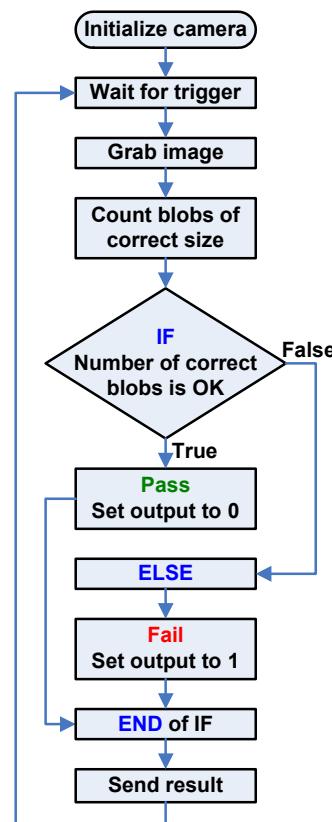
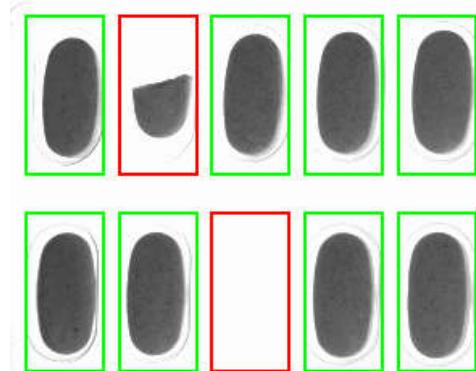
Example

A blister pack needs inspection before the metal foil is glued on to seal the pack. The task is to inspect each blister for pill presence and correctness. If any of the blisters is faulty, the camera shall conclude a fail and the result is communicated to reject the blister pack.

A backlight is used to enhance the pill contours.

The application can be solved either by pixel counting, pattern matching, blob analysis, or edge finders, depending on accuracy needs and cycle time requirements.

The **flow diagram** below illustrates the idea of the blister pack program when a blob analysis approach is used.



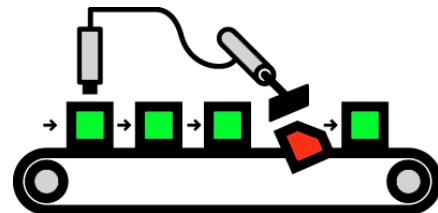
Flow diagram that describes the camera program for a blister pack application.

6**Communication**

A machine vision system can make measurements and draw conclusions, but not take actions by itself. Therefore its results need to be communicated to the system in control of the process.

There are a number of different ways to communicate a result, for example:

1. Digital I/Os
2. Serial bus
3. Ethernet network.



In factory automation systems, the result is used to control an action, for example a rejector arm, a sorting mechanism, or the movement of a robot.

The hardware wiring of the communication channel can either be a **direct connection** or via a **network**.

6.1 Digital I/O

A **digital I/O** (input/output) signal is the simplest form of communicating a result or receiving input information. A single signal can be used output one of two possible states, for example good/bad or true/false. Multiple I/Os can be used to classify more combinations. In industrial environments, the levels of a digital I/O are typically 0 (GND) and **24V**.

6.2 Serial Communication

Serial communication is used for transmitting complex results, for example dimensional measures, position coordinates, or read strings.

Serial bus is the term for the hardware communication channel. It transmits sequences of bits, i.e. ones and zeros one by one. The communication can be of three kinds:

1. Simplex, one-way communication
2. Half duplex, two-way communication, but only one direction at the time
3. Full duplex, two-way communication.

The speed of data transfer is called **baud rate**, which indicates the number of symbols per second. A typical value of the baud rate is 9600.

There are many kinds of serial buses, where the most common in machine vision are:

1. RS232 (Recommended Standard 232). Can be connected to the comport on a PC.
2. RS485 (Recommended Standard 485)
3. USB (Universal Serial Bus).

6.3 Protocols

A **protocol** is a pre-defined format of communicating data from one device to another. The protocol is comparable to language in human communication.

When a camera needs to communicate a result, for example to a **PLC** system (Programmable Logic Controller) or a robot, the result must be sent with a protocol that is recognized by the receiver. Common PLC protocols are:

1. EtherNet/IP (Allen Bradley, Rockwell, Omron)
2. MODbus (Modicon)
3. DeviceNet (Rockwell)
4. Profibus, Profinet (Siemens)
5. FINS (Omron)
6. IDA (Schneider).

6.4 Networks

Many cameras operate on **networks**. A network is a communication system that connects two or more devices.

Network communication can be described by hierachic **layers** according to the **OSI Reference Model** (Open System Interconnect), from the lowest hardware level to communication between two software applications:

1. Application layer (software-to-software communication).
2. Presentation layer (encryption)
3. Session layer (interhost communication)
4. Transport layer (TCP, assures reliable delivery of data)
5. Network layer (IP address)
6. Data link layer (MAC address)
7. Physical layer (wiring and bit-level communication)

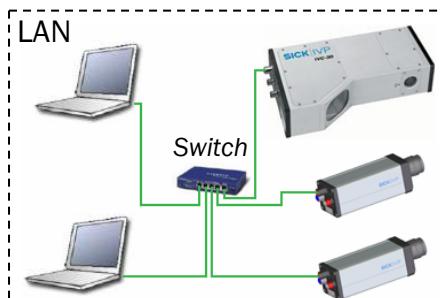
6.4.1 Ethernet

Ethernet is the most common networking technology. The Ethernet standard defines the communication on the physical and data link levels.

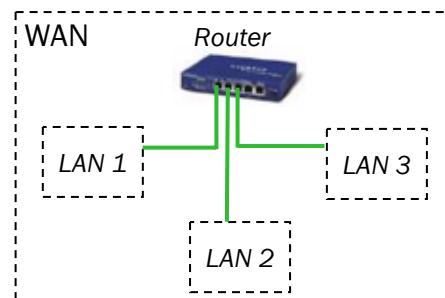
Ethernet exists in different data transfer speeds: 10, 100, and 1000 Megabits/s. The fastest of the three speeds is also known as GigE or **Gigabit Ethernet**.

6.4.2 LAN and WAN

An Ethernet **LAN** (Local Area Network) connects various devices via a **switch**. When two or more LANs are connected into a wider network via a **router**, they become a **WAN** (Wide Area Network).



Example of a LAN (Local Area Network).



Example of a WAN (Wide Area Network), connecting multiple LANs.

Additional information on Ethernet LAN communication in practice is available in the Appendix.

7

Vision Solution Principles

Choosing and implementing machine vision technology involves the following questions:

1. Is vision needed to do the job?
2. Is there a financial incentive for investing in machine vision?
3. Is the application solvable with vision?
4. Which vision technology should be used?
5. What does a typical vision project look like?
6. What problems might be encountered along the way?

7.1 Standard Sensors

Vision is a powerful and interesting technology, but far from always the best solution. It is important to keep in mind the vast possibilities with standard sensors and also the option of combining cameras with standard sensors.

A simple solution that works is a preferable solution.



7.2 Vision Qualifier

When assessing the suitability of an application to be solved by machine vision, there are certain economic and technical key issues to consider.

7.2.1 Investment Incentive

Vision systems are seldom off-the-shelf products ready for plug-and-play installation, more often they should be considered as project **investments**. The reason is that vision solutions almost always involve some level of programming and experimenting before the application is robust and operational.



The first step is thus to determine if there is a **financial incentive** or **justification** for an investment. There are four main incentives for this investment:

1. Reduced cost of **labor**: Manual labor is often more costly than vision systems.
2. Increase in production **yield**: The percentage of the produced products that are judged as good-enough to be sold.
3. Improved and more even product **quality**: The actual quality of the sold products through more accurate inspections. Even a skilled inspector can get tired and let through a defect product after some hours of work.
4. Increase in **production speed**: The output can be increased wherever manual inspections are a bottleneck in the production.

The price of the vision system should be put in perspective of the investment incentive, i.e. the combined effect of reduced labor and increase in yield, quality, and production speed.

A rule of thumb is that the total cost of low-volume applications are approximately twice the hardware price, including the cost of **integration**.

Once the financial incentive is defined, a feasibility study can be considered.

7.2.2 Application Solvability

Image quality

Having good contrast conditions and high-enough resolution is essential. A poor or variable image quality can sometimes be compensated for by the use of algorithms, but developing them is costly and robustness is an issue. In general, it is worthwhile to strive towards the best-possible image quality before going on to the image processing.

A key factor in building a robust vision application is to obtain good **repeatability of object representation** regarding:

1. Illumination
2. Object location and rotation

There are methods to deal with variations in these factors, but in general less variability gives a more robust application. The optimal situation is a shrouded inspection station with constant illumination where the object has a fixed position and rotation relative to the camera.

Image Processing Algorithms

Having a good image addresses the first half of the solution. The next step is to apply image processing **algorithms** or **tools** to do the actual analysis. Some vision systems are delivered with a ready-to-use software or **toolbox**, whereas others need third-party algorithms or even custom development of algorithms. This can have a heavy impact on the project budget.

7.3 Vision Project Parts

Once the application has qualified as a viable vision project, the phases that follow are: Feasibility study, investment, implementation, and acceptance testing.

7.3.1 Feasibility Study

The purpose of a **feasibility study** is to determine if the problem can be solved with vision or not. In the feasibility **report**, there is information about the application, what parts have been solved and how, and which problems and challenges can be expected if the application becomes a project.



The feasibility study should either reach **proof-of-concept** (meaning "Yes, we can solve the application"), identify why the application is not solvable, or state which further information is needed before proof-of-concept can be reached.

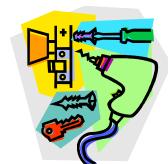
7.3.2 Investment

Once the feasibility study is complete, it's time for the **investment decision** and the **project definition**. The project definition should contain a full description of what the vision system shall do and how it will perform. A procedure for acceptance testing should be included.



7.3.3 Implementation

The **implementation** is the practical work of building the system. The contents and extent of the implementation phase may vary from a partial to a complete solution.



Implementation is often called **integration**. When a company provides integration services, they are referred to as a **system integrator**. When vision is the integrator's main business area, they are referred to as a **vision integrator**.

7.3.4 Commissioning and Acceptance Testing

Once the implementation phase is completed, it is time for **commissioning** of the system, or handing it over to the customer. A part of the commissioning is an **acceptance test** according to the procedure described in the project definition.



The acceptance test description contains clear conditions of customer expectations of the system. If the test is passed, the system is considered to be completed or delivered.

7.4 Application Solving Method

The general method for solving a vision application consists of the following steps: Defining the task, choosing hardware, choosing image processing tools, defining a result output, and testing the application. Some iteration is usually needed before a final solution is reached.

7.4.1 Defining the Task

Defining the task is essentially to describe exactly what the vision system shall do, which performance is expected, and under which circumstances.

It is instrumental to have **samples** and knowledge about the industrial site where the system will be located. The collection of samples needs to be representative for the full object variation, for example including good, bad, and limit cases.

In defining the task, it is important to decide how the inspected features can be parameterized to reach the desired final result.

7.4.2 Choosing Hardware

The steps for selecting system hardware are:

1. The type of object and inspection determine the choice of **camera technology**.
2. The object size and positioning requirements determine the **FOV**.
3. The smallest feature to be detected and the FOV size determine the **resolution**.
4. The FOV and the object distance determine the lens' **focal length** (See the Appendix for explanations and example calculations of needed focal length.)
5. The type of inspections and the object's environment determine the choice of **lighting**.
6. The type of result to be delivered (digital I/O, Ethernet, etc) determines the choice of **cables** and other **accessories**.
7. The choice of camera and lighting determines the **mounting mechanics**.

The above list is just a rough outline. Arriving at a well-defined solution requires practical experience and hands-on testing.

When standard hardware is not sufficient for the task, a **customization** might be needed.

7.4.3 Choosing Image Processing Tools

In choosing the processing tools for a certain application, there are often a number of possibilities and combinations. How to choose the right one? This requires practical experience and knowledge about the available tools for the camera system at hand.

7.4.4 Defining a Result Output

The next step is to define how to **communicate the result**, for example to a PLC, a database, or a sorting machine. The most common output in machine vision is pass/fail.

7.4.5 Testing the Application

The application is not finished until it has been tested, debugged, and pushed to its limits. This means that the system function must be tested for normal cases as well as a number of less frequent but possible cases, for example:

1. Ambient light fluctuations, reflections, sunshine through a window, etc.
2. Close to the acceptance limit for good and bad.
3. The extremes of accepted object movement in the FOV.

7.5 Challenges

During the feasibility study and implementation of a vision application there are some challenges that are more common than others.

This section treats typical bottlenecks and pitfalls in vision projects.



7.5.1 Defining Requirements

It can be a challenge to define the task so that all involved parties have the same expectations of system performance. The customer has the perspective and terminology of his or her industry, and so does the vision supplier. Communication between both parties may require that each share their knowledge. To formalize clear **acceptance test conditions** is a good way of communicating the expectations of the system.

7.5.2 Performance

The **cycle time** can become a critical factor in the choice of camera system and algorithms when objects are inspected at a high frequency. This situation is typical for the packaging and pharmaceutical industries.

Accuracy is the repeatability of measurements as compared to a reference value or position (measure applications). Accuracy is described in more detail in the Appendix.

Success rate is the system's reliability in terms of **false OKs** and **false rejects** (inspect and identify applications). A false OK is when a faulty object is wrongly classified as OK, and a false reject is when an OK object is wrongly classified as false. It is often important to distinguish between the two aspects, since the consequences of each can be totally different in the production line.

7.5.3 System Flexibility

Building a vision system that performs one task in a constant environment can be easy. However, the system's **complexity** can increase significantly when it shall inspect variable objects in a variable environment.

Worth keeping in mind is that objects that are very similar in the mind of their producer can be totally different from a vision perspective. It is common to expect that since the vision system inspects object A with such success, it must also be able to inspect objects B and C with the same setup since they are "so similar".

7.5.4 Object Presentation Repeatability

The **object presentation** is the object's appearance in the image, including position, rotation, and illumination. With high repeatability in the image, the application solving can be easy. On the other hand, the application solving may become difficult or even impossible for the very same object if its presentation is arbitrary.

For example, **rotation invariance** (360 degree rotation tolerance) in a 2D application is more demanding on the processing than a non-rotating object. In 3D, rotation invariance might not even be possible for some objects because of occluded features.

7.5.5 Mechanics and Environment

Although a vision system can be a technically optimal solution, sometimes there is not enough **mounting space**. Then one must consider an alternative solution or redesign of the machine.

Some industries have environments with **heat**, **vibrations**, **dust**, and **humidity** concerns. Such conditions can have undesirable side-effects: Reduced hardware performance, life, and deteriorated image quality (blur). Information about the hardware's ability to withstand such conditions is found in its technical specifications.

8**Appendix****A Lens Selection**

Selecting a lens for a straight-forward application is a three-step procedure:

1. Measure the object dimensions, or the maximum area in which the object can be located.
2. Measure the object distance.
3. Calculate the needed focal length.

In high-accuracy applications or under other special circumstances, selecting an appropriate lens may require special knowledge and considerations. The section about telecentric lenses below describes such an example.

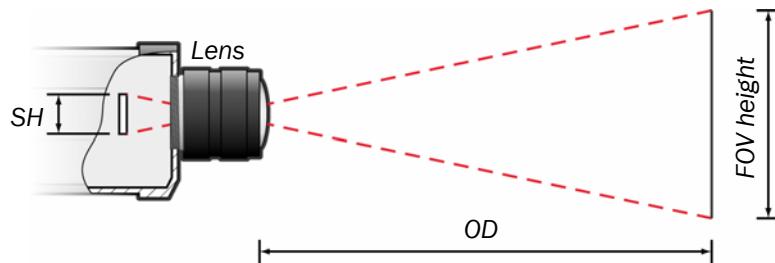
Calculation of Focal Length

A suitable focal length can be calculated with the following formula:

$$\text{FocalLength} = \frac{\text{SensorHeight} \cdot \text{OD}}{\text{FOVheight}},$$

where OD is the object distance.

The formula is a simplified model intended for practical use, not for exact calculations.



The figure above can just as well represent the sensor and FOV widths. The important thing is to be consistent in the formula by using only heights or only widths.

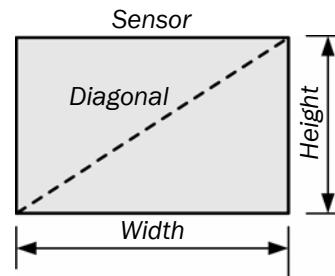
The sensor size is often measured diagonally in inches. SensorHeight in the formula above refers to the vertical height, which thus needs to be calculated from the diagonal.

Example

The focal length for a certain application needs to be calculated.

Known information:

- Camera: IVC-2D VGA (640x480)
- Sensor size 1/3 inch → diagonal 8.5 mm
- Object distance = 500 mm
- FOV height = 100 mm.



To use the above formula, the sensor height needs to be calculated first. The needed calculation principles are the Pythagorean Theorem to calculate the sensor diagonal in pixels, and the theorem about uniform triangles to find the sensor height in mm.

Calculations:

$$1. \ SensorDiagonal = \sqrt{640^2 + 480^2} = 800 \ pix$$

$$2. \frac{8.5 \ mm}{800 \ pix} = \frac{SensorHeight}{480 \ pix} \rightarrow SensorHeight = 5.1 \ mm$$

$$3. \ FocalLength = \frac{5.1 \cdot 500}{100} \approx 25 \ mm$$

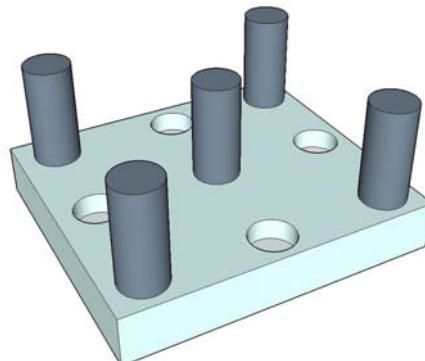
Telecentric Lens

Normal lenses give a slightly distorted image because of the view angle. **Telecentric lenses** can be used to reduce or eliminate such effects. The optics in a telecentric lens causes all light beams enter the lens in parallel. This has two practical consequences:

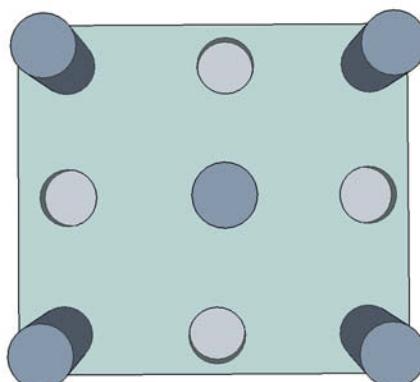
1. The lens diameter needs to be at least as large as the object's size.
2. The depth of field is deeper than for standard lenses.

Telecentric lenses are mostly used in high-accuracy measurement applications.

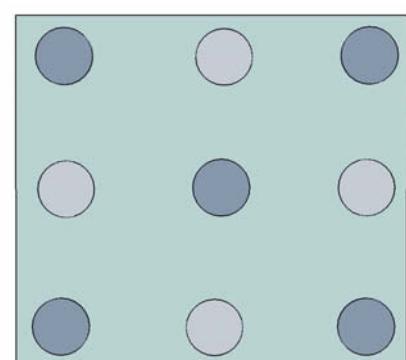
Example



Object with cylinders and holes. The task is to measure their diameters.



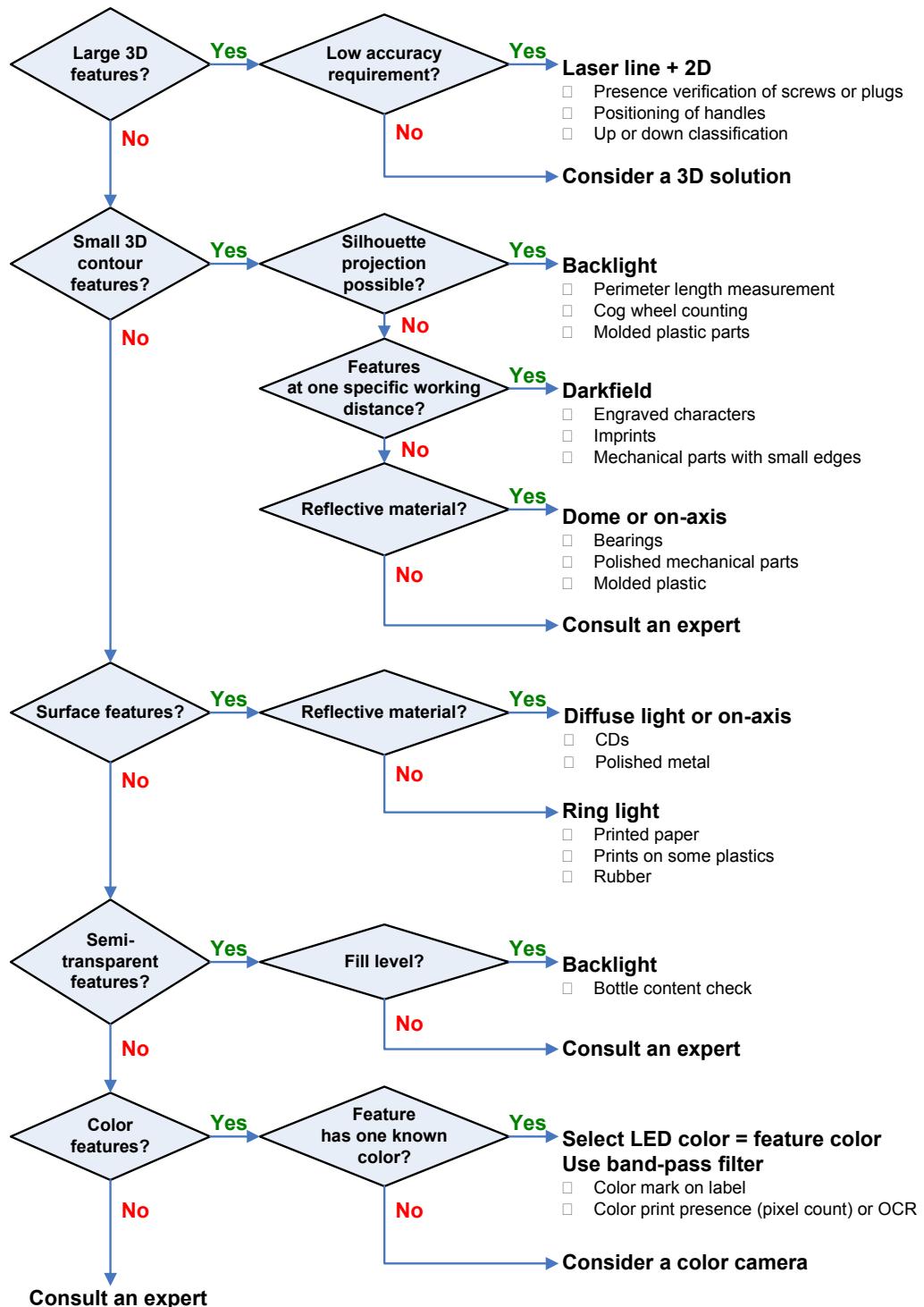
Cylinders as viewed through a normal lens from above.



Cylinders as viewed through a telecentric lens from above.

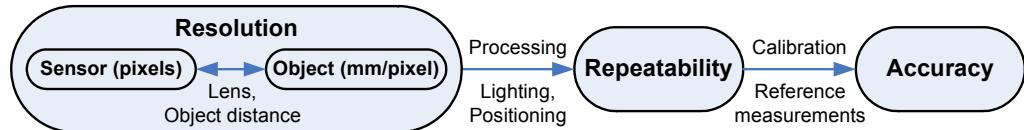
B Lighting Selection

Selecting the best lighting for a particular application can be tricky. The flow diagram below is a rough guideline for typical cases.



C Resolution, Repeatability, and Accuracy

Resolution, repeatability, and accuracy are terms sometimes used interchangeably, which is incorrect and can lead to false expectations on a vision system. The terms are connected in the following way:



Sensor resolution is the number of pixels on the sensor, for example VGA (640 x 480).

The sensor resolution together with the focal length and object distance (i.e. the FOV) determine the **object resolution**, which is the physical dimension on the object that corresponds to one pixel on the sensor, for example 0.5 mm/pixel.

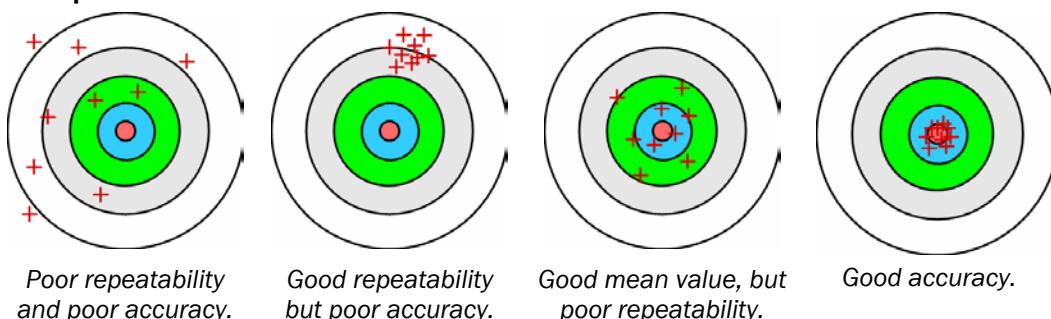
The object resolution together with the lighting, analysis method, object presentation etc determines the **repeatability** of a measurement. The repeatability is defined by the **standard deviation** (σ , 3σ , or 6σ) from the **mean** value of a result when measured over a number of inspections.

Accuracy is the reliability of a result as compared with a “true” value, or **reference value**. Accuracy is defined in terms of the standard deviation (σ , 3σ , or 6σ) from the reference value, usually referred to as the **measurement error**.

If the repeatability of a measurement is good there is a fair chance that a **calibration** procedure will also give good accuracy. In the simplest case the calibration is just subtracting an offset, such as the function of the zero button (tare) of an electronic scale. In vision applications it can be a much more complex procedure, involving specially designed **calibration objects**, lookup tables, interpolation algorithms, etc. A consequence of calibrating against a reference is that the vision system won't be more accurate than the reference method.

If measurements of the same features change systematically over time, the system is **drifting** and needs **recalibration**.

Example



Confidence Level

A vision system is claimed to measure the diameter of a steel rod with 0.1 mm accuracy. What does 0.1 mm mean? The **confidence level** at which you can trust the system depends on how many standard deviations (σ) the stated accuracy value refers to:

- **σ** : You can be **68%** sure that the measured value is within 0.1 mm of the truth.
- **3σ** : You can be **99.7%** sure, i.e. on average 3 measurements out of 1000 have a larger error than 0.1 mm.
- **6σ** : You can be **99.9997%** sure, i.e. on average 3 measurements out of 1.000.000 have a larger error than 0.1 mm.

Thus a more correct way of stating the accuracy of the diameter measurement is to include the confidence level: "...diameter of a steel rod with 0.1 mm (3σ) accuracy".

Which confidence level to choose is application and industry dependent, though 3σ is usually a reasonable choice for acceptance test conditions.

D Motion Blur Calculation

Motion blur occurs when the exposure time is long enough to allow the object to move a noticeable distance during this time.



Sharp image.



Motion blur because of too long exposure time relative to object speed.

Motion blur can be calculated in physical dimensions b (mm) or in pixels p :

$$b = t \cdot v$$

$$p = \frac{d}{r}$$

where

b = physical distance (mm)

p = number of pixels

t = exposure time (ms)

r = object resolution (mm/pixel)

v = speed (mm/ms or m/s)

Motion blur needs to be considered in applications where the conveyor moves fast or when small features need to be detected accurately.

Motion blur is reduced by increasing the light intensity, for example by strobing, and decreasing the exposure time.

E IP Classification

IP Classification is a system for telling how resistant a device is to **dust** and **water**. This is specified by a two-digit number, for example IP65. The first digit in the IP class tells how well-protected the device is against dust, and the second digit indicates the water resistance.

IP65
First digit, dust Second digit, water

The table below specifies the classifications commonly used for camera systems. Grayed-out numbers are included for reference.

First digit	Meaning	Second digit	Meaning
4	Protected against 1.0 mm objects	4	Protected against splashing water
5	Dust protected	5	Protected against water jets
6	Dust tight	6	Protected against heavy seas
		7	Protected against immersion down to 1 m, water tight for a short while
		8	Protected against submersion longer periods of time, water tight
		9	Protected against high pressure water jets (80-100 bar jets at 100 mm distance, 80 °C) or hot steam

It is important to note that the system is not more resistant than its "weakest link". It is of no use to buy a camera of IP67 if the cable connectors are IP53.

Unusually harsh environments, for example where there is an explosion risk, require additional classification systems.

F Ethernet LAN Communication

Ethernet communication on a **LAN** (Local Area Network) is based on the standard **UDP** and **TCP/IP** protocols. (IP, or Internet Protocol, has nothing to do with the IP classification treated in the previous appendix.) The UDP protocol sends data on the network without any confirmation that the data reached its final destination, whereas the TCP/IP protocol ensures that no data is lost during the transfer by acknowledge and resend functionality.

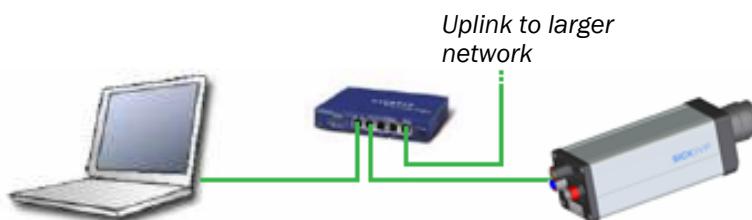
IP Address and DHCP

To enable communication between two devices on a LAN, they need unique **IP addresses** (Internet Protocol). This works in a similar way to a normal telephone line, which also needs to have a unique number for identification. The IP address consists of four 8-bit numbers (0-255) separated by decimal points, for example 132.15.243.5.

The IP address is either **static** or **dynamic**. Static means that it is constant and dynamic means that it is set by an external **DHCP server** (Dynamic Host Configuration Protocol). The task of the DHCP server is to give each device on the network a unique IP address each time it is switched on, so **IP collisions** are avoided.



*Simplest possible LAN: Direct connection between computer and camera.
No DHCP server requires static IP addresses.*



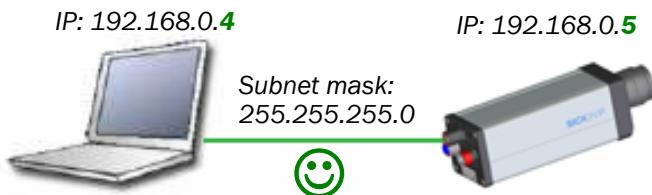
*Standard LAN: Connection via a switch. IP addresses can be dynamic
if a DHCP server is connected to the network via the uplink port.
If not, both the computer and the camera need static IP addresses.*

Subnet Mask

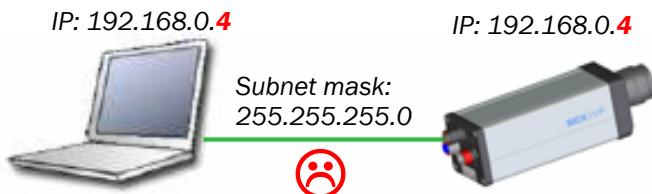
Ethernet TCP/IP, the protocol for the Internet, enables communication with devices all around the world. For efficient communication, the local network needs to be restricted. This is partly achieved by a software function called **subnet mask**, which puts constraints on each 8-bit number in the IP address:

1. **0** means: Allow any number.
2. 1-254 means: Advanced use, don't bother at this stage.
3. **255** means: Both devices' numbers at this position in the IP address must be identical.

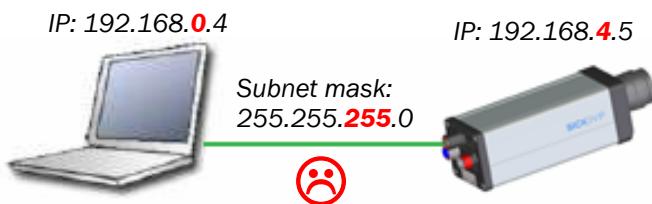
A common subnet mask is 255.255.255.0, which requires the first three numbers to be identical on both devices, whereas the last can be anything but identical.

Example: Networking with Static IP

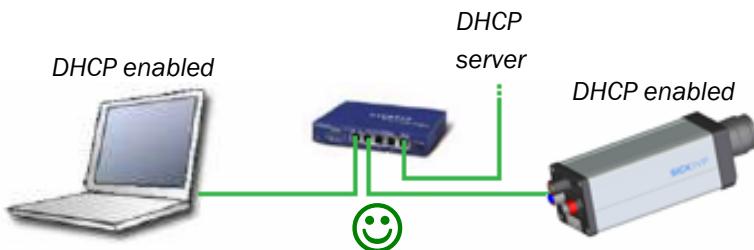
Example 1: Static IP addresses are unique and the communication works.



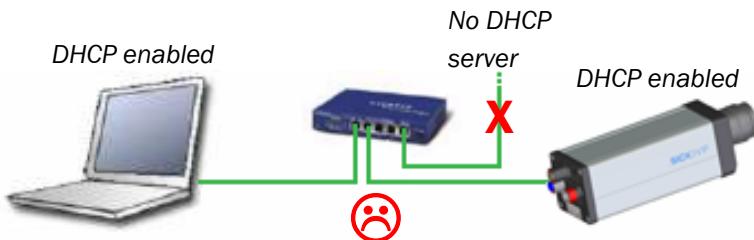
Example 2: Identical IP addresses cause a collision and the communication doesn't work.



Example 3: The subnet mask requires the first three numbers in the devices' IP addresses to be identical, which is not the case and the communication doesn't work.

Example: Networking with Dynamic IP (DHCP)

Example 4. DHCP is enabled in both PC and camera, and their IP addresses are provided by the DHCP server on the network.



Example 5. DHCP is enabled in both PC and camera, but no IP addresses are provided without a DHCP server and the communication doesn't work.