Robot Vision

Bruce Batchelor Ralf Hack Paul Whealan

June 3, 1997

Contents

1 Introduction

2 Vision

2.1 Human Vision

Everyone is an expert in vision [3]. Human vision is the most dominant sense for all day life. It is the result of several million years of evolution resulting in a extrodinary simlicity and efficiency. Although seeing is associated with the eyes, human "image sensors" are only one part of a complex biological system.

The most powerful part about human vision is our life expertise and common knowledge used. It enables us to extract and understand what we want to see from what we actually receive from the environment. This knowledge is a part so powerful, that it allows us to see even without eyes, by using imagination and insight.

The link between seeing and understanding is one of the major challenges in computer based vision. For a computational process an image is comparable to book in a foreign language to the reader. The reader may receive all images necessary to read but without semantical and syntactical knowledge about the content all information remains "unreadable". In the same way, a computer has to draw from the programmers knowledge to extract useful information from images.

The knowledge is implicitly implemented throughout a vision system. It is the underlying structure everywhere: in the setup for the image aquisitation, in the image processing and in the interpretation of the results.

However, even with a lifetime of experiences, images can fool the human perception, which has been demonstrated remarkably by Escher [**?**]. An artificial observer applying parts of our knowledge faces the same drawback. The image aquired may be right but if the system has no knowledge about its rightness, "does not see the big picture", then the results and conclusions reached may be unnoticed entirely wrong. This is often the case, when computer aided vision is used in conditions which have not been considered during the design. A famous example is the factory floor, which is exposed to sunshine only at a certain time of year and this additional light does distort the image aquisitation and therefore the production grinds to a halt for no apparent reasons.

2.2 Computer Vision

Computational and logical supremeity of computers over the ability of mankind always triggered imagination and fears about computers as supreme being, uncorruptable decision maker, performing better in all the disciplinces. Originally, databases, expert systems and vision systems may have been designed for such a purpose.

However, until now every single system built with those high aims did not succeed in all but its special domain. As soon as common knowledge, common sense and life experience are involved, those system fail miserably.

Nevertheless, many useful services have been implemented with computers and research is being carried on to improve those systems further. Computers may be soon in the position to drive cars safer, faster and cheaper than a human driver. The communication with road and traffic data sources may enable them to navigate and find the best routes available. All of which is being done by every experienced driver, and a few drivers may be achieve some excellence. Using their knowlege maybe provitable to everyone.

This is where computer vision has its stronghold. Computer vision tries to find the complex concepts of human vision and to establish the underlying concepts and methods. It also expands those concepts to overcome known limitations and expands the boundaries of our experiences about our environment.

Work in this research discipline has been carried out to built computer able to see like us. Most of the research into this area is aimed to understand how human vision is being done and to learn to overcome human visions limitations.

In due course, several subdisciplines established themselves: two dimensional and three dimensional vision, motion recognition and tracking, signal processing or artificial intelligence based vision.

2.3 Robotics

As with computer vision, most of us played with the idea to automise repeatative task mechanically in some sort of artificial worker. Robotics started off with this aim and the human abilities as a physiological model. The idea was to built a manipulator which interacts with the environment for a multitute of purposes.

As with vision, human like flexibility was certainly the goal, but for many applications not necessarily required. Accuracy and relyability in repeatative tasks proofed to be just as useful. This shaped this new engineering discipline with the aim to do build an efficient solution. This solution should perform a task with methods and devices available, integrate this task into a factory environment and automate it up to a level necessary for cost efficient production.

Specialist roboters are now doing their work everywhere. Transport robots navigate and schedule transportations tasks on factory floors. Machining robots use tools and manipulators to perform manufacturing tasks by welding, glueing, cutting, graving, positioning or handling of objects with high precision, relyability and accuracy.

According to their tasks, robots with althogether different function and layout have been developped. So emulate SCARA robots still human physiology but Gantry robots are able to work in cartesian coordinate systems. This make them a good companion to the current camera technology, usually based on cartesian coordinates.

The design of task for robots requires extensive knowlege about robotics, cybernetics and mechanics as well as process control and human interfacing, each discipline is a part in its own right but altogether more than the sum of its parts.

2.4 Machine Vision

Depending on the point of view machine vision is claimed to be an extension to robotics, computer vision, system engineering or applicated physics. However, just as Robotics is more than controlled motion, machine vision is more than computer vision.

In contrast to Computer Vision, the domain of Machine Vision aims to develop new solutions for new applications. The challenge in machine vision is not the understanding of images but the integration of vision into an application, especially designed to be used for a certain predefined purpose.

So does Machine Vision generally use images, but applications may dispose of the need for visual images and use ultrasonic, x-ray, Eddy current or taktile sources which might proof a lot faster and cost efficient than any other solution. This article discusses methods and principles in Machine Vision using visual images.

Picture @@graphic(vision path) does give an overview about a typical machine vision system. Three regions of vision can be distiguished: Early vision, low level and high level vision. Additionally, there are the areas of positioning, handling and transport, process control, user control and automation.

3 Vision system

This section discusses the different vision related regions of machine vision: physics based vision and image aquisation, low level vision and high level vision. Physics based vision is the part of each machine vision system, which is used to capture the image before digitizing it. Low level vision is the fast logical and mathematical part of an image processing system, usually used to enhance and optimize the image for further processing. Literature refers to this level also as early vision or signal processing. Recognition and information extraction is the area of high level image processing. It creates a abstract image representation for further processing, usually connected with recognition and evaluation processes.

3.1 Image aquisitation and physical based vision

An image is the result of three components interacting with each other: light source, object and light sensor. The light source is emitting energy, which bounces off or is transmitted through a number of surfaces. A small amount of this energy is then scattered into the general direction of the sensor. A wavelength and intensity subset is then detected by the image sensor and accumulated over a specific time period. The sensor quantizises the received energy according to its spatial position. This representation is then called the image.

Every single light beam detected represents what is visible to the sensor. Since most of the light does not reach the sensor, some information is being lost in the image formation process.

Physical based vision tailors this behaviour to extract the required information. By deliberately restricting the incoming information content, the image evaluation is being simplified and improved in its relyability, speed and robustness. Examples and design studies can be found at [2].

Hence, physic based vision is a very versatile part of an image processing system, which is being done before an image has been digitized and processed by the computer.

The mathematical basis is the bidirectional reflection distribution function (BRDF). The BRDF is modelling the surface reflection according incidence, observer and surface orientation and intensity and reflection properties of the surface [1]. Using the BRDF and a reflection model of the object, either the objects micro and macrostructure, the observer or illumination can be derived. Essentailly, as in ray tracing, every single light beam is modelled and followed.

The models can be distinguished into physical models and geometrical models. Physical models are based on the electromagnetic wave theory, whereas geometrical models use only a subset of the electromagnetic spectrum and its behaviour based on empirical studies and appearance [5]. The most common geometrical model in use is the Lambertian model where each incident light beam results into an unidirectional reflection. The more elaborated model Torrance and Sparrow defines a rough surface as a collection of planar micro-facets with their orientation normally

distribution. As an example for a physical model, the Beckmann-Spizzichino model based on the Maxwell equations. Being more general and complex, this model also includes effects like polarization.

Machine Vision is based on the principle of selecting the required information by surpressing and elimination of the redundant and unnecessary part. It can be argued that Machine Vision is therefore a lossy compression technique. The first step of data reduction takes place in physics based vision. The overall quality of each vision system is dependant on the ability to accurately extract the information required. Information which is not available as this stage cannot be handled and evaluated in the following processing. Too much information at this point unnecessary complicates further processing.

Therefore, the first and foremost design goal is to choose the early imaging architecture to enhance the feature of interest and at the same time reduce the noise to aquire the right data from the environment under observation.

3.2 Low level image processing

At this stage, the image has been aquired and needs to be processed to further improve visibility of the features of interest. It is typically stored in a raw form with spatial and intensity resolution according to the image digitization. The aim of this stage is to isolate and identify the feature of interest. The resulting image after this stage is typically a binary image with a shape or edge based representation to be further processed by the following high level image processing.

Noise is any information content in the image which is of no use to the task. It is typically caused by the choosen early imaging process and the devices related to it. Typical noise are unwanted parts of the object and light source, distortions and reflections from other sources, structure and modulations from the light source, etc. Most of the processing at this stage will reduce the amount of noise or unwanted information to achieve the maximum signal to noise ratio.

Low level image processing uses fast and simple mathematical and logical techniques to reduce noise. These techniques are choosen according to spatial and intensity distribution of the feature, the noise and early vision architecture. As a typical first step, the image is being evaluated statistically to obtain and compensation on individual grey level variation. This is being done to improve robustness of the following processes.

The processing continues by applying a variety of operations, selecting the feature from the background according to its typical characteristics. Most of the applied operatios are selected empircally. This is a rather crucial and sensitive area, which is usually the cause of errors in later stages, some author describe it righly as art [?]. A more detailed describtion of the used image processing operations, especially the low level image processing can be found in chapter 5.2. One successful strategy to erase unnecessary details is to extract the unwanted parts and execute subtract or exclusive or operation between the original image and the extracted parts to eliminate them.

3.3 High level image processing

After the low level image processing stage, the next step is the high level processing. At this stage, the image is semantically analysed to extract the required information. In many cases, this information consists of geometrical or statistical data extraction for measurement purposes. In other cases, the information from the image needs to be converted into a different representation, e.g. depth map from structured light images.

A major part of high level image processing is the area of pattern matching. A selected reference pattern is compared against the image to find content or absence of features. Because of the enormous complexity of the process, this requires a fair amount of knowledge about the parts to be successful in time, if at all.

To segment an image into the objects represented, usually edges together with the details about the type of objects is used to distiguish the different objects in the image and to process them individually. Edge segmentation is also used for tracking and navigation purposes. The low level image processing has to make sure, to enhance the areas of large gradients. The Hough transform, section ??, is very powerful in detecting lines or circles. Other techniques use "bugs", which follow pixels if they are arrange to a uninterrupted pattern. (some reference required here).

Intuitivaly, objects are classified according to their shape. This is the way morphology works. Similar to the pattern matching, the image is processed using shapes. The two main operations are erosion and dilation. Erosion removes and shrinks shapes of the selected sort, whereas dilation does the opposite. It expands white pixels to the shape given.

After the high level vision process, a vision system needs to export the results for further processing, statistical evaluation or process control. This is part of the systems engineering task, which is highly individual to the current use.

4 Lighting Design and Physics Based Vision

Light is the vehicle being used to collect information from the environment. The advantage of using light is that the object is under low energenic influences which do not interfere with the objects natural behaviour. It also is a fast and relyable information transmitter and is mostly harmless to and easily controllable by a human observer.

As mentioned before, the most important part in most applications is to receive the right light from the feature of interest. Every other light has to be filtered out and increases the complexity of the task. And obviously, light which is not received by the sensor can be the missing information.

This chapter gives an overview over light sources, lighting configurations and light sensors to optimize the Physics based vision and image aquisitation and therefore reduce the complexity and resources of later stages.

4.1 Light sources

The light emitted from a light source has three typical properties: spatial distribution, temporal distribution and temperature or colour. The spatial distribution is the projected light pattern from the light source to the object including polarisation. The temporal distribution is the behaviour of the light intensity over time, usually strobed or continuous light or homogenity of lasers. The light colour is dependant of the material used to produce light energy.

In many cases, the light source is combined with a optical system, e.g. collector, mirrors, filters, lenses or even fibre optics to enhance desired lighting qualities.

Common light emittors are currently classified into three broad categories: gas discharge sources, semiconductor sources and heat based sources.

4.1.1 Semiconductor Sources

Semiconductor sources, also known as LEDs (light emitting diode) consist of two different donated semiconductor materials. The electrical energy which transmits

through the contact zone between the two materials looses some of their energy in form of light. Due to the fact that the materials are very pure, the light energy emitted has well defined and narrow color properties. And since the light is generated in the contact zone only, the emission direction is restricted.

LED sources are generally fairly dimm sources since the energy efficiency is only about Eff=0.5%..5%. By choosing the semiconductor material and donation, the colours can be selected from the whole range from infrared to green.

This technique allows for very cheap, compact, robust, fast switching and low powered light sources with an almost unlimited life span. With the small colour bandwith, LED are almost monochromatic light sources. To improve on the drawback of the dim lighting, Multi-LED arrays are available in different geometrical arrangements. To achieve stroboscopical effects the LED's are being strobed which has also the advantage of boosting the brightness for the required short period of time.

4.1.2 Head based souces

Heat based sources use the natural light emmisions of hot objects, usually generated by electrical resistors which glow with strong electrical currents. The most common version is the vaccuum light bulb which consist of a Wolfram wire in a vaccuum. More advanced sources use a halogen atmosphere which enhances the emission process and therefore increase the overall efficiency.

Due to the nature of the physical processes behind it, the light distribution is generally unidirectional. Also, the colour distribution is broad ranging from infrared to white with a strong infrared component. The color is dependant of the material and the amount of energy supplied. Due to the physical stress on the material whilst powering up, the life span is relatively short. The light bulbs have to be shielded from rapid temperature influences, shock and contact.

This light source is commonly used because of the high energy output, low costs and availability. Most of the light sources are equipped with a optical system to focus the light energy and additionally filter infrared light to save objects form the immense heat generated. There is a large variety of geometrical shapes of the wire from points to lines, which can be used to customize the the illumination characteristics.

4.1.3 Gas discharge sources

Gas discharge sources use a gas atmosphere which is pumped by electrons to discharge ultraviolet light. This light is then converted to different colours by floureszent materials. After starting by a high energy pulse gas discharge sources are very energy efficient.

This type of source is ideal for geometrical extended light sources like area or ring lighting. Modern lamps can be as compact as large light bulbs with the integrated starting electronic, which also allows lamps to work on low voltage power supplies as in cars or electrical torches.

4.1.4 Other light sources

A special type of discharge sources are the plasma lamps. Between two special electrodes a plasma is ignited which heats the electrodes to emitt light and melt and burn in the process. Although very bright, these sources are short lived and need constant maintanance. Some developments are being done on improving the plasma emission to safe the electrodes and achive a better efficiency.

Although the area of image aquisitation is usually connected to visible or near infrared light, a large area of applications are based on UV, Infrared, CT or X-Ray

Lighting	Surface properties	Transparency
Front	visible by reflection	generally invisible
Back	generally invisible	visible
Darkfield	visible by reflection	visible by defraction
Brightfield	visible by reflection	visible by defraction
Unidirectional	visible by shading	visible by defraction
Structured	required to be homogen	visible by defraction
Polarized	visible by polarization	visible by polarization

Table 1: Lighting principles and application principles on solid or transparent bodies

systems. Some areas like ultrasonic and even taktile area sensors can be used to generate image of the environment to be inspected. The largest area is the non contact volume inspection in medical or technical areas.

4.1.5 Fibre Optics

A variant to the conventional optical systems to achieve variable source illumination patterns are light sources combined with optical fibres. Based on glas or quartz minerals those fibre transmit the photonic energy very efficient from one end to the other.

Since the fibres are flexible, they can be arranged like MultiLED lamps in almost any geometrical shape. This constellation compensates for many of the disadvantages like heat emission and shock sensitiveness of the halogen lamps by seperating source from the illumination point.

The transmission quality is highly sensitive to the light color. The best quality can be achieved using IR light. The fibres have a low frequency filtering effect on modulated and monochromatic signals because of signal delays due to multidiffraction, which makes them also sensitive to movements.

4.2 Lighting design issues

commentconfiguration overview

As shown above, illumination is a central part in all machine vision applications. B.G.Bachelor [2] collected over hundered typical configurations and applications of early vision configurations. Literature empirical [**?**] distinguishes seven principal lighting configurations, table 1, which work either on reflection or defraction. Whereas reflection causes a light direction change on solid bodies, defraction changes light directions whilst transferring through bodies. Table 2 distinguishes the principle property for detection surface properties.

4.2.1 Front lighting

This is the most common lighting configuration in use for human vision and computer aided visual inspection. A light source illuminates an object and the reflected light is then collected by the image sensor.

The light source and sensor enclose an angle less than $\phi_{ref} = 90@deg$, with the effect that the sensor does not receive the light directly from the source. The light received is either reflected from the object or background. Usually a low reflectance background is selected. This light then carries information about the surface and shape properties of the object.

To inspect either of the properties, the other property has to be known. An accurate shape inspection depends on an accurate surface reflection model. In the same way, a unknown shape can distored surface reflectance, especially with

Lighting	Microstructure inspection	Macrostructure inspection	3D inspection
Front	by reflection	by shading	by photometry, stereo
Back	invisible	silhoutte inspection	generally not possible
Darkfield	generally visible	visible, see chapter 4.2.3	possible
Brightfield	generally visible	visible, see chapter 4.2.3	possible
Unidirectional	visible by shading	generally invisible	generally not possible
Structured	required to be homogen	by triangulation	by triangulation
Polarized	visible by polarization	visible by polarization	visible by polarization [?]

Table 2: Principles for inspection surface properties of solid bodies

spherical objects. Problems usually encountered are cause by interfering of the shape into the surface inspection by shadowing, occlusions and obstacles.

This configuration does collect the maximum information about the object. It is also used to extract shape information using structured illumination and triangulation as mentions in chapter 4.2.5.

Mirror, glossy or specular reflection can render this method unusable due to the danger overloading of the sensor. The large information content of the images received leads to a high redundancy and in some circumstances to cancellation of information. Another typical problems is caused by foreign light sources. Light from outside the system, e.g. sunlight, factory illlumination, reflected by the object can have disturbing effects on the early vision task. Therefore, some effort has to be made to isolate the setup from outside influences, especially the very strong sun light.

4.2.2 Backlight illumination

In contrast to the front light illumination, the light source and the light sensor oppose each other. Light emitted from the source will fall onto the light sensor area. The object will diffract or deflect light. Hence the information carrier is the missing part of the illumination.

The light source, object and light sensor are positioned along the optical axis of the system. The light source is usually an area illumination source, which needs to be larger than the feature under inspection. The object is poseing in a way that all features under observation are obstacles for the light and the other features merge into the shadow area.

For solid or opak objects, information about the surface is lost in the process. This fact leads to a robust and simple technique for shape extraction of small and medium sized flat objects. Images aquired with this technique are usually already binarized used by morphology and shape inspection algorithms. Particular useful is this configuration with object inspection on carrier tapes, e.g. integrated cirucuit, etc. With the carrier semitranspartent, the current position of the element is easily detectable.

Diffuse backlight usually causes some smearing of the edges of thick objects. This can be reduced using telecentrical illumination in which all light rays travel

into the same direction. Another cause for blurring and smear effects is bright light overloading the camera, which can easily be handled by adjusting the aperture accordingly. The light source has to emitt homogen light because the light pattern is received as background. For efficiency reasons this background must be easily separable from the dark object.

4.2.3 Bright or Dark field illumination

To achieve surface structures like encravings or relief prints the material, usual metal based, experiences minute changes in its height structure. Light incidence under a low angle results in a wide range of scattered light beams covering almost the whole range of $\phi_{ref} = 180@deg$. By selecting a observer position at a specific angle above the surface, particular reflection can be choosen to selectively inspect a structure according to its height and slope of the relief.

To achieve this, the sensor and light source usually enclose a reflection angle between $\phi_{ref} = 90@deg$ to $\phi_{ref} = 180@deg$. The light source emitts a highly directional light. As with the front illumination, the sensed light is received by reflection only. The light sources are usually choosen to be laser or light stripes.

For the bright field illumination, the camera is adjusted to receive most of the reflected light. The absence of received light is the information carried. This is typically used for diffuse or transparent surfaces. Transparent surfaces diffract the light internally. This light can then be used to measure thinkness and homogenity of the translucent material, see also [2].

The dark field illumination uses the light deflected from the often specular light stream for inspection. This is usually used for gravings and measuring of surface roughness or fabric quality.

With cameras at different positions, minute radians and reliefs can be detected and measured. The typical glossy reflectance of metal is highly directional. Therefore, a camera position outside the main reflection angle enables the measurement of surface roughness and structure. With a additional spatial filter, some areas of the camera can be blinded of to receive all but the main relecteion.

This method is very sensitive to the position of the object. Minute changes can falsify the image and render it useless. It is also dependant on the material of the object used. With the increase of the roughness of the surface, the sensor will be overloaded.

4.2.4 Unidirectional Illumination

The unidirectional lighting configuration uses an allaround light source which has no preferred direction of incidence. It is comparable to the optical phenomens inside a fog. All shadow and specular reflection is lost.

It is usually generated by indirect lighting by a diffuse white hemnisphere illuminated along the perimeter. The object has to be shielded from the direct light.

This is advantageous for glinting and voluminous objects with holes which tend to produce unwanted shadowing effects.

But it also imposes thermal stress since the light source needs to be fairly bright. The construction needs to contain the whole object. The observation has to be made through holes and is therefore fixed in its position.

4.2.5 Structural illumination

Using structural illumination a special light pattern is projected onto a surface. This pattern is often used to extract 3D information from the surface of the object. By using a light stripe pattern the surface shape can be extracted using triangulation methods.

The structural illumination is based on the foreground illumination principle. The pattern projector and the sensor are positioned with a defined angle between $0@deg \leq \phi_{proj} \leq 90@deg$. By measuring the shift along the light stripes, the height is then proportional to the product of the shift and the sine of ϕ_{proj}, see also [?]. To detect the light stripes, the surface pattern should not interfere with the shape detection, e.g. no specular reflection and reasonable smooth (Lambertian).

Another use for structural illumination is the Moiree effect, when two identical pattern projected onto a surface visualise minute changes in height or reflection properties. [more about moiree, please]

Structural illumination is a robust method of detecting objects which have a multicolored surface. The fact that only the existence of connected light stripes is required, it is a fairly robust technique. The major advantage the use in 3D scanning. Partial illumination of objects is a very fast way of presence/absence inspection since there is hardly an image processing overhead required. Structural illumination is also used in conjunction with line scan cameras, which only require illumination of the object part under observation.

Spherical shapes or considerable surface slopes and rought soufaces can diffuse or interrupt the light pattern which will decrease measurement accuracy. Another difficulty are objects with different reflection abilites. This might change the visibility of the light stripes considerable. Some of the reflection dynamic can be compensated for low contrast areas by adjusting aperture which can cause a merging or blurring of other areas. Areas which have a low local incidence angle need to be treated for blurring, since the surface roughness broadens the pattern roughly accoring to the tangens of the incidence angle.

4.2.6 Polarisation Effects

Reflected or transmitted light does almost never maintain its emittion polarisation. The polarisation ratio between the two perpendicular components changes with the incindence angle. By filtering the one component, the irradiance can be filtered out complete at this angle. The angle where the reflection is polarized in one direction only is called Brewster angle. For transparent dielectrical media, the angle is around $\phi_{brewser} = 56@deg$. The polarisation is detectable by filtering the light through a optical polarisation filter. According to the filter orientation the horizontal or vertical part of the light wave can pass through.

The polarisation measurement is an add on to other illumination principles. Filters can be applied either to the light source or before the light sensor. Some light sources, like lasers and LED do have a prefered polarisation, whilst gas discharge or heat based sources experience random or circular polarisation. [Do cameras have a detection polarisation preference ?]. Polarisation is also in some cases dependant on the reflection angle. This behavour is used to measure minute relief changes [?].

Polarisation filtering is especially useful for mixed glinting and lambertian surfaces. A lambertian or diffuse reflection does not produce a harmonic polarisation, each reflected ray has a different polarisation ratio, whereas the specular reflection usually experiences a harmonic polarisation. With a polarisation filter, the lambertian reflection is reduced by a maximum $P_{red} = 50\%$ in any filter orientation, whereas the specular reflection can be eliminted completely for the Brewster angle.

Other uses may be found in inspection of translucency of transparent material, which do not have a strong influence on the visibility but in many cases changes the polarisation.

With standard cameras and lighting devices, most applications can be upgraded and improved using filters. Polarisation can be used to inspect minute surface changes, e.g. stress measuring of tools.

Light filters have a decreasing effect on signal strength, which has to be compensated. Note that in many cases, a different lighting configurations is more robust and relyable than a glinting filter. However inspecting flat and polished metal surfaces can make such a filter necessary.

4.2.7 Light sensors

4.2.8 Semiconductor Sensors

The two commonly available light sensors or cameras are either semiconductor based or electron tube based.

There are four different principles of semiconductor cameras [?]. Self Scanned Photodiodes (SSPD), Charge Coupled Devices (CCD), Charge Injection Devices (CID) and Charge Coupled Photodiodes (CCPD).

SSPD and CCPD sensors use simple photodiode elements as pixel sensor. These elements either generate the signal converting the received light energy (generator princip) or act as charge bucket. The SSPD uses the generator prinzip and has a linear output signal to the light intensity. The power generation produces unwanted heat which interferes with the linearity. Therefore those elements need to be cooled, usually by a peltier element. The CCPD sensors use the bucket principle [true or false]

CCD and CID sensors are based on field effect technology, comparable with field effect transistors. This principle uses capacitive effects and can therefore work powerless for their operation. The CCD charge a pixel beforehand and light discharges it according to the light energy. (?) CID (?)

To extract the status of each pixel to form an image, two different principles are in use. The charge couple principle moves the charge of each pixel into a analog shift register and destroying the status of the read pixel in this process. The injection method accesses each pixel by measuring the charge of each pixel sequentially.

The CCD and CCPD sensors are based on the charge couple method. The charge of each a pixel row or column is shifted into an analog shift register in parallel. The shift register is then clocked to read the status sequentially. This method has a fast but fixed readout sequence. Each pixel in a row has exactly the same exposure time. Due to the limitations of the analog shift register, the signal experiences distortions like blooming and image smearing because of crosstalk between the register elements. Therefore this technology needs brighter images for a good signal to noise ratio. The advantage of this technology is in the simplicity of the structure.

The charge injection method uses a multiplexer to access one pixel at a time. This allows for random access readout. Due to the high isolation between pixels, the distortions by blooming and smearing are neglegible. The charge remains intact for integration and exposure time control, and needs to be erased in a separate cycle.

Both principle use a fast analog to digital converter (ADC), although in theory the analog signal can be used for PAL or NTSC output signals. The use of the ADC allows for efficient internal compensation to achieve auto iris effects. The major advantage is the fixed spatial resolution which allows for high repeatability of the aquisitation. Other advantages: the semiconuctor technology typcial long live span and almost no ageing process.

Even with no light input, each pixel produces a small output signal which is highly depential to the temparture. Hence the raw pixel signal is being subtracted to assign black to zero output. Standard cameras are designed with a 4:3 ratio in the image size. In some cases, the pixel shape ration reflects this. Since most image processing operations are defined on square pixels sizes, the image needs to be adjusted to reflect the asymetrical rations for accurate measurements.

To compensate for changing overall illumination to achieve an optimal brightness resolution, the exposure time is usually adjustable. The basic principle is to feed the minimum and maximum output signal levels into the time span generator.

4.2.9 Electron Tubes

Electron tube sensors use silizium eletrical storage coating on the tube head with a very large time constant. A scanning beam charges this capacitor. Incident light does discharge it again. The scanning beam does recharge the capacitor area once each frame and at the same time registers the necessary current. This current is a measure for the incident light energy.

The advantage of electron tube sensors is the highly cusomizable sensitive frequency range. Additionally, the sensitive area is variable. By changing the scanning area on the tube inside, interesting areas can be zoomed without loosing resolution.

However, the size, weight, power consumption and mechanical sensitivity of the tube are disadvantages which limits its usefulness, especially with cheaper semiconductor cameras and the progress in adjustable lens systems. Other problems is repeatability and geometrical distortions especially near strong magnetic fields (electrical engines, power supplies, etc.). The semiconductor capacitor is very sensitive to overload and constant exposures, burn in's, therefore it has a limited live span.

4.2.10 Camera systems

[standards PAL,NTSC, HDTV, PAL plus]

4.2.11 Camera optics

5 Image Processing

5.1 Image representation

The image representation is the format in which the data from a light sensor is stored for save keeping or further processing. There are hundreds of formats in use, either propretary or public domain, each with advantages and disadvantages in respect to storage size, pixel accessability, abstraction level and usage. Generally, one can distiguish two major representations: DCG and vector based or pixel based.

DCG or vector based representation make use of two or three dimensional abstract objects, e.g. lines, cubes, rectangles to build up the image of an object. Each feature is then stored with its drawing command and all its parameters and styles. The drawing process is done by interpretation of the commands and executing it.

The requirements on storage space is dependant on the complexity of an image. The image size of a vector based images is almost linear dependant on the amount of details represented. Each feature of a represented object is usually represented by its edges.

Vector or DCG based objects are widely used for CAD and virtual reality. One very common type are postscript images. Generally, it is very efficient for man made objects.

The DCG representation does allow for platform and graphic device independant storage. The independancy of this format does give the drawing process the ability of adjusting the image to the possible output format, e.g. color optimization, customization of resolution and view port. The view port is of special interest for 3D imaging. In this case, the viewer may adjust the point of the camera and the view port interactively. Some interpreters, e.g. Postscript, even allow for programs

Image Type	representation method	typical size
Binary image	each pixel is either black or white	1 bit[1]
Gray level image	each pixel classifies a grey level in a predefined range[2]	1,2,4 bytes
floating point	each pixel is represented by a floating point	4,8,.. bytes
complex images	each pixel is represented by a tuple of two values	twice the numbers of the above

Table 3: Monochrome pixel images storge sizes

Image Type	representation method[3]	typical size
color indexed	each pixel is a color index of a color lookup table	1 byte
real color images	each pixel is a color representation	3,4 bytes

Table 4: Color pixel images storge sizes

to be stored into the image, which might be used for simulation and moving images. The format is very efficient in storage size for man made drawings and objects, in general for objects which can be represented by prototypes.

Since the drawing is a combined interpretation and displaying process, this system does require computing resources. It also shifts the control of the displaying process to the end device. The output image layout is therefore less predictable. This format is less efficient for free form drawing or non man made objects because the storage size depends on the ability of representing the object algorithmic in drawing commands. This cannot be done efficient for handwriting, plants or just pixel based representations.

The other format of representation is the pixel based representation. For the pixel representation the object is being sampled in either one or two dimensions. This is done by mapping it onto a plane. This plane gets quantisized into rectangle or square samples of some known dimension. Once aquired, the image is fixed like a photograph and can be stored into an array of numbers. The type of numbers is to be defined with the aquisitation process, because information lost here cannot be recovered later. Table 3 gives an overview over used representations.

The storage size is dependant on the resolution required. Without compression, the image size is independant from the resolution and sampled area size. Storage space not used by an object is used by the background.

Simple images contain large monoton areas, usually background or surfaces. To reduce storage size, such images can be compressed by the usual mathematical compression methods with some success. In this case, the increase of image complexity does increase the image size, since the mathematical redundance decreases.

For uncompressed images, increasing the complexity inside the image does not change the overall image size. This representation is compatible to most output devices. Therefore, the displaying process does require no complex processing. In most cases, resizing and color adjustment is all that is required, which makes it a fast and efficient displaying format.

Since the image does not reflect the original object, there is no method of rescaling and zooming into details. The resolution is fixed from the point of aquisitation and can get only worse.

5.2 Image processing operations

Image processing operations apply a mathematical or logical function on a set of input pixels.

operation	functionality	comment
thresholding	binarizes the image	pixel value is inside a value range
pixel transformation	simple mathematical transformation	apply mathematical operation e.g. add value to, log all values
histogramm distribution functions	extract statistics or apply histogram transforms	get the grey level and equalize it
coordinate transform	cartesian to radial shift, rotate, warp image	move pixel around into a coordinate representation

Table 5: Principle Monadic Operations

In contrast to image processing operations are graphical operations. Whereas image processing cannot increase the overall information content of an image, graphical operations add information or transform the shape or orientation of the image. Examples for graphical operations are drawing, area fill routines as well as crop, resize or rotation operations.

Defined by their mathematical complexity, there are seven main classes of image processing operators. Monadic operations apply pixel to pixel modifications, Dyadic operations apply pixel to pixel modifications by combining two or more images. Local operations apply image processing operator on a window of the source image(s) which is being shifted over all pixel. Morphological operations are a subset of local operations which modify a particular pixel region usually applying logical local operations. Histogram operations apply statistical evaluation onto the image and image transformation use the input image to calculate the output pixel, e.g. Fourier or Hough transformations.

5.2.1 Monadic operations

Operations of the $g_{(i,j)} = M(f_{(i,j)})$ are called monadic operations. Each resulting pixel corresponds to one input pixel. The typical order of these operations is $\varsigma(n)$, where n is the amount of input pixels being accessed. Each pixel is processed individually, independant of its position. Table ?? shows the commonly used operations of this class.

Monadic operations allow for a fast processing implementation using look up tables on grey level or color index images. These tables can be calculated in advance. The complexity of the algorithm is in most cases linear.

The used processing method is actually a signal stream processing without taking the image content and pixel position into consideration.

5.2.2 Dyadic operations

Dyadic operators use two or more images in conjunction. The mathematical operation is usually described by $g_{(i,j)} = D(f1_{(i,j)}, f2_{(i,j)}, .., fn_{(i,j)})$. Each result pixel corresponds to two or more input pixels. The order of this algorithm is $O(N * n)$, where N is the number of input images and n the amount of pixel per images. The ususal requirement is that the input images are of the same size and representation.

Dyadic operations allow for image integrations to reduce noise and spontaneouse influences. Special care has to be take so that the alignment of the obejct is consistent overall source images. Another useful properties is to mask parts of the image by logical operations.

Since dyadic operations combine several image, the result usually requires a larger pixel value storage size than the source images to avoid overrun errors. There-

operation	functionality	comment
arithmetical	combines two images	integrates or diffrentiates two images
logical	combines two images	applies maskings

Table 6: Principle dyadic operations

operation	functionality	comment
filter based	apply a filter function onthe local gradient	integrates or diffrentiates successive pixel val
logical based	apply a rank filter on the local neighbourhood	take the min or max pixel value of the windo

Table 7: Principle local operators

fore most operations are in the largest possible storage size and normalised afterwards. Since the pixel position needs to be consistent over all source images, these images need to have the same size and resolution.

5.2.3 Local operators

Local operators use only a subset of the source image at a time, usually a square neighbourhood of the current pixel under observation. The mathematical describtion is $g_{(i,j)} = L(f_{(n1,m1)}, f_{(n2,m2)}, .., f_{(nx,my)})$. The order of this algorithm is $O(n * m)$ where n is the amount of pixel of the input image and m the amount of pixel in the window for small windows. With large windows, many of the border lines are being ignored because of the window size, reducing the number of pixels. For rank filters, the order increases due to the sorting routines used each window up to $O(n * m^2)$. Common window sizes are 3x3, 5x5 or larger. Odd numbers allow the destination pixel to be in the center of the image, which is not a necessary requirement. These windows are also known as kernels.

Dyadic operations need to have an accurate alignment, whereas local can operate on slightly misaligned images as well. Due to the spatial components, the operations can detect gradients in every direction and enhance or compensate for it. Typical application is noise filtering, shape enhancement or local shape detection.

Special consideration has to be taken while processing near the image boarders. In most cases the access of pixel outside the image is avoided by ignoring border near image rows and columns. Other processing methods define the outside area to a predefined background color, usually black. The processing time is exponentially dependant on the window size, since every pixel needs to be accessed multiple times while the window is shifted along.

5.2.4 Morphological operations

Morphological operations are applied to modify or detect shapes in images. It seems to be a natural and easy way to think of shapes in the algorithmic design of applications. Morphology can be explained as part of pattern matching, where a correlation between a given shape and the to be inspected image needs to be extracted.

As mentioned above, these operators are defined as local operators applying logical function on their neighbourhood. Morphological operations are also known as iconic image processing because of their shape sensitiviness.

Operations are usually performed on binary images, where the object is white and the background is black. There are two types of operation, erosion and dilation. Erosion will shrink a particular shape if found to a point or remove it altogether, whereas dilation will explode each point found to the operator shape.

operation	functionality	comment
erosion	remove occurences of operator pattern	also known as opening operation e.g. removes small connections
dilation	expand each pixel to operator pattern	also known as closing operationе.g. merges regions.
pattern matching	compare pattern and shapes	find defined shape forms

Table 8: Principal Morphological Operations

operation	functionality	comment
fourier transformation	spatial to frequency transformation	
cosine / sine transformation	spatial to discreet cosine transformation	compression techniques
hough transformation	cartesian to r/theta transformation	

Table 9: Principal Image Transforms

Generally, morphological operations are defined to work on binary images. To extend their functionality to grey level images, methods of the fuzzy logic can be employed. A dilation function is then defined as a local maximum operations and erosion function become local minium operations.

Applications are found in different areas. To detect cracks in images, a combination of two dilation to delete small cracks, two erosions to undo the size changes without restoring the cracks. The resulting image is then compared by an exclusive-or to find the removed cracks.

The blob labelling operations is useful to index each blob with a particular grey level. This allows the processing to be selectively applied onto a particular blob. The blob labelling is explained in [?] and works according to the area fill algorithm.

Since this processing works selectively on particular shapes, they provide a precise tool.

The complexity does require considerable computation resources to be effective. To recognize the pattern, the operator has to have the same size and orientation as the pattern to be recognized. This might be compensated by normalising scale, position and orientation of the image with the penalty of the use of more computational resources. It is also noted, that pattern matching can be implemented using the fourier domain.

5.2.5 Image transformation

Other image processing operation transform images into different representations. Each resulting pixel is the result of a calculation of all input pixels. The typical order is $O(n^2)$, with n as the number of pixels. For special functions, like the FFT, the order can be as good as $O(n \log_n)$.

Some operations move the pixel positions, like cartesian to radial coordinate or r/Θ-transformation, translations, rotations, warping. Other operations change the representations from spatial to frequency domain, e.g. fourier or DCT transformation.

Operations like the fourier transformation are very powerful, because they translate frequency processing into the much simpler spatial processing. Particular pow-

operation	functionality	comment
contrast enhancement	expand to full storage range	compensate illumination variation
histogram modification	change histogram function	equalise histogram image enhancement
local histogram modification	change histogram of a local area	enhance pattern

Table 10: Principal Histogramm Operations

erful examples are scale independant image comparisons and pattern matching. Sine and Cosine transformations are used in many cases, to get advantages of the FT togehter with the fast processing of spatial operationts.

The hough transforms allow for geometrical interpretation of images. For the standard hough transform each point in the hough image is corresponding to a line in the input image. Other transforms are sensitive to circles.

Because of their high complexity, these operations are usually very slow. Additonally, since most operations require floating point operations, hardware implementations are complex and expensive.

5.2.6 Histogram operations

Histogram operations perform a statistical evaluation of the input image. The results are being used to improve image quality and feature visibility. The order is typical $O(n)$ with n as the number of input pixels. Since many histogram functions are being applied again, the resulting order is $O(2n)$.

Histogram modification routines are very robust and usually used to improve images after aquisitation by eliminating minor lighting and object reflection changes. Local histogramm equalisation is particular usefule to enhance patterns, e.g. using local histogram equalisation. This is usually a two stage process, stage one extracts the histogram then the histogram gets modified and a lookup table is generated. The second stage maps the pixel values form the input image to the output image. [?] uses this method to estimate an optimal threshold for images with monomodal histograms [++].

The complete image has to be scanned to extract the histogram. Therefore this technology needs large buffers when used with for fast pipline processors. Every histogram operations usually requires considerable amount of memory for calculation buffers. The buffer size is to be estimated by $Memreq = N_{pix} * N_{grey}$ for each buffer with N_{pix} the number of pixels in the image and N_{grey} the number of grey levels a pixel can have.

5.3 Implementation issues

Modern personal computers are powerful enough to be able to process most image processing operations with acceptable speed for most purposes.

Since more and more system can be easily equipped with image aquisitation hardware, the need for dedicated hardware to perform image processing operation decreases rapidly. Even real time image processing is possible.

Monadic and some dyadic operations can implemented in modern image aquisitation hardware by using available look up tables for color and grey level modifications. With a new morphological algorithm like the SKIPSIM technique [?], real time morphology and binary local opeartions are now possible and has been done [?].

Dedicated high speed hardware is mainly used for demanding high level or vector based image processing. The area of frequency domain computing, e.g. fourier imaging and high quality rendering and modelling are typical domains. Computers archectures used are pipeline processors, e.g. [??] for sequential pixel map processing.

In many cases, only a particular task or functionality has to be done with high speed. Designed for such purposes various hardware extension systems are available, with variing degrees of processing power. Some boards use typical signal processing equipment which can be programmed to execute programs on a stand alone basis. The host computer is merely the interface, see also [?]. Examples are being found in the area of transputer, neuronal nets or pipeline processors. Other systems are frame grabber with some signal processing abilites.

6 System Design

Some image processing applications are designed around the devices at hand or the computers available or profitable to use. A more advantagable approach is to start off with the objects and features to be inspected.

After all, one of the first questions before deciding for a vision based system should be: Is a vision system the best method for a particular application or is there any other way to resolve the task with simple sensoric or different process design?

Image processing systems are integrated devices. An image processing application usually requires other components. Objects need to be accurately positioned and transferred from and to the inspection area. The image processor needs some knowlege about the process for the inspection and the inspection results have to be supplied in time and synchron to other processes, used and stored on other system components. To increase robustness and accuracy, the image processor, especially cameras and lighting are shielded against outside influences, like heat, radiation, vibrations, electrical noise, foreign bodies and people. Error conditions have to be detected and handled.

All these aspects have to be carefully considered before designing the image processor and the system around it. Before any ideas for the image processing application can be search for, the designer has to check a whole range of system specifications which narrow the range of designs down. Those specifications are:

- Speed: At what speed are results needed for the operation. Important bottlenecks have to be identified, like image aquisitation speed or frame/line rates, image transfer speeds (computer bus speed). Current standards allow for about 15 frames per second without image processing time. Early vision operations can be performed with one or two frames delay.

- Spatial Resolution: What is the physical resolution of the object feature to be inspected. The standard resolution of cameras is about 1000 times smaller as the maximum object feature size which can to be simultaniously aquired. Requirements exceeding this resolution need special expensive systems, designs and imaging devices.

- Accuracy: The maximum real accuracy of an image processing system should never exceed the size of one pixel for spatial accuracy and 0.2% of the maximul grey levels. It is possible to push the spatial limits mathematically by a factor of 10, but only at computational costs. The dynamic range of 2^{10} of CCD image sensors may be exceeded with advanced cameras technology.

- Flexibility: Specification about the range of objects to be inspected should be considered as minimum recommendations to the designers. Because of the nature of colliding expertises in image processing and specific application area, common sense expectations on the customer side, the image processor needs to be easy maintainable and customizable. Especially, cheap equipment costs are usually punished with high maintanance costs whilst and after installation. The designer should listen very carefully to extract all the process requirements, especially those build on common sense.

As an empirical rule, the above specifications should be carefully estimated and systems should be designed with a healthy safety tolarance. Technological improvements in the applications might make it necessary to increase the specifications. Especially inspection systems experience the first redesign requirement withing two years (???) of use.

6.1 Calibration Issues

To allow for robot vision interaction, both system have to be calibrated to a compatible coordinate system.

6.1.1 Camera calibration

Cameras do experience two kinds of imaging inaccuracies: spatial and signal errors which distort the signal size and digitisation position of individual pixels[4]. Additionally, the image sensor itself may experience movements by temperature and time. For a discussion of the errors induced by the optical system, the reader is refered to the literature [?].

The following example demonstrates the physical and quantisational sensitivity of a visual measurment process. Current technology uses image sensor with around 1024×1024 pixels. The sensor size is about $A_{sens} = 5 \times 5mm$. By using a field of view of about $A_{view} = 500 \times 500mm$, each pixel digitizes an area of about $0.5 \times 0.5mm$. Considering a healthy savety distance, errors exceeding about $E_{rat} = 10\%$ have to be considered harmful. In our example this would amount on the object side in positioning errors of $E_{Opos} = 0.05mm$ and on the sensor side with about $E_{Spos} = 0.5\mu m$. Considering E_{Opos} in conjunction to positioning equipment or E_{Spos} on the background of vibration, it becomes clear of the scope of errors induced by the sensor itself.

Spatial errors are related to the one or two dimensional resolution of the image. [?] Several sensor technology design related errors can be identified:

- The sensitive pixels area is because of compatibility reasons for standard video signal norming rectangular, with a dimension ration of 4:3. This ratio may vary further from sensor to sensor and between different rows and columns. Because of drifts in the horizontal synchronisation, the framegrabber might shift rows against each other whilst digitisation.

- The output video or digital signal is usually a sequential readout of the intensity value of a row. Small jitter effects shift these rows agains each other. This is particular important for images taken from several cameras at the same time, e.g. stereo imaging.

- Because standard video norms images are usually transmitted in half images, the vertical synchronisation between the two half images may vary and cause the framegrabber to shift digitisation.

Signal errors occur during energy conversion from photonic energy to electrical energy and energy transport [4].

- Semiconductor materials experience a significant dark current. This current is usually compensated by reference signal elements internally. However, there is a small individual pixel dark current left due to inhomogenities in temperature and cristal structure.
- Processing errors in the fabrication lead to site to site variations in sensitivity and background noise generation.
- Because only a certain color can induce photocurrents, the signal experiences a signal color dependant sensitivity.

In addition to the local imaging errors, the camera also experiences positioning errors.

- For measurement purposes, the image sensor needs to be positioned perpendicular to the object under oberservation. A tilt of the camera will produce perspective distortions, proportional to the square cosine of the error angle horizontal and vertical.
- The image sensor is usually soldered onto a circuit boards. Under temperature influences boards tend to warp and the image sensor will therefore move.

To calibrate a camera system, the standard calibration model is used which is based on nine variables: 3 positioning variables, 3 orientation angles, focal length and resolution in both dimensions of the imaging plane. Several approaches are being used to estimate all the values. For a complete discussion, see [?]. The standard method uses a grid of calibration markers. The standard camera model is then used to calculate all the variants in advance.

This method is fast and relyable for standard purposes and short live cycles of the calibration.

This model does not compensate for dynamic errors, e.g. temperature drift. Extensions of this model are made to allow for variable focal length, [?].

comment camera/robot calibration A simple way of calibration for both camera and robot is to mark several positions inside the visible area by placing an easy detectable object. With both the camera and robot coordinates known in cartesian coordinates, the appropriate calibration matrix is easily obtainable.

References

[1] *Robot Vision*. MIT Press, 1986.

[2] B.G.Batchelor. Lighting advisor.

[3] B.G.Batchelor and Paul.F.Whealn. Machine vision systems: Proverbs, principles, prejudices and priorities. *Evening Workshop in Machine Vision Applications, Architectures, and Systems Integration III, Proc. SPIE 2347, Boston USA*, page 374.

[4] Glenn E. Healey and Raghava Kondepudy. Radiometric ccd camera calibration and noise estimation. *PAMI*, 16:267.

[5] Shree K. Nayar, Katsushi Ikeuchi, and Takeo Kanade. Surface reflection: Physical and geometrical perspectives. *PAMI*, 13:611.