Chapter 4

Recording and Displaying of Images

4.1 I - Introduction

This chapter discusses *image formation*, i.e. everything that preceeds the actual image processing by a computer. Light detection equipment (photodetectors and cameras), optical instruments (lenses, mirrors, and glass fibers), light sources (lamps and illumination techniques), and image displays are discussed.

A good vision system design will start with the careful examination of appropriate choices for the above components. Not only should they be made compatible, but also can a good choice facilitate further processing dramatically. The robustness, reliability, and speed of a vision system will often rely more on the issues discussed in this chapter – although they may seem to be side issues – rather than anything else.

4.2 I - Cathode Ray Tubes

4.2.1 Principles

For reproducing brightness patterns from the electrical signals into which they were transformed by a camera, so-called *cathode ray tubes* (CRTs for short) are still the most widely used devices. A remark in order here is that flat displays are increasing their share of the market. Three flat display technologies are used: liquid crystal displays (LCD) (especially with active matrix addressing, ALCD), electroluminiscent displays (ELD) and plasma displays (PD). The flat display market is currently dominated by LCDs, with an estimated share of 60% for 1995. Some important problems remain: yield and addressing problems for ALCDs, colour problems for ELDs and PDs. Plasma gas screen emit light by energizing neon gas sealed between two sheets of glass. The energized gas gets into a plasma state, where it emits ultraviolet rays. They in turn activate phosphors as in CRTs, which generate red, green, and blue light. A discussion is outside the scope of this introduction.

The global structure of CRTs is shown in fig. 4.1. CRTs have a phosphor screen which lights up in accordance with the intensity of the impinging electron beam. The

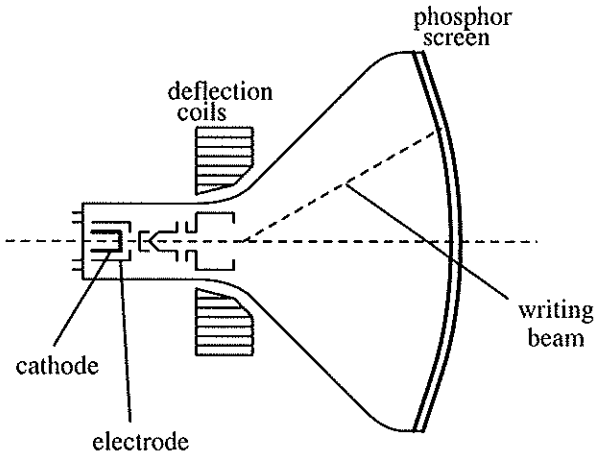


Figure 4.1: *Black & white picture tube with electron beam deflection system.*

electron beam, generated by the cathode, is deflected by magnetic fields produced by the currents in the horizontal and vertical deflection coils. The intensity of the electron beam is influenced by the voltage across the control electrode. The pattern is divided into a number of lines which are explored from left to right and from top to bottom. For this the scanning beam is deflected horizontally and vertically, writing a line raster. Synchronizing pulses are transmitted to ensure that the writing beams cover the correct picture elements. The scanning process for video is discussed further in the next section.

4.2.2 The video standard

In order to provide the desired viewing quality, the physiological characteristics of the eye have to be considered. In particular, the discrete character, both spatial and temporal, of displays should be invisible to the viewer. The separate lines and phosphor elements must not be discernible and flickering has to be avoided. Under normal conditions the limit angle separating the lines must not surpass $\alpha = 1.5'$ as shown in fig. 4.2. Good viewing conditions for creating a realistic sensation have total vertical viewing angles of 20 degrees or more. This corresponds to a distance from the screen that is about three times the picture height (or less). On the other hand, this is the minimum viewing distance not causing fatigue of the eyes. Optimal aspect ratios (i.e. width/heigth) are in the range 5 : 3 to 2 : 1.

The quality of the reproduced picture is determined by the resolution, which is the better the higher the number of lines with a minimum being required to ensure that the raster isn't disturbing the viewer. The viewing distance D was chosen to be five times the picture height H, $\frac{D}{H} = 5$. Although not quite in agreement with the above given optima, this choice was also to take account of technological limitations at the early days of television. At this distance the line picture should just be no longer visible. From

$$tan(\alpha) = \frac{H/L}{D} \sim \alpha$$

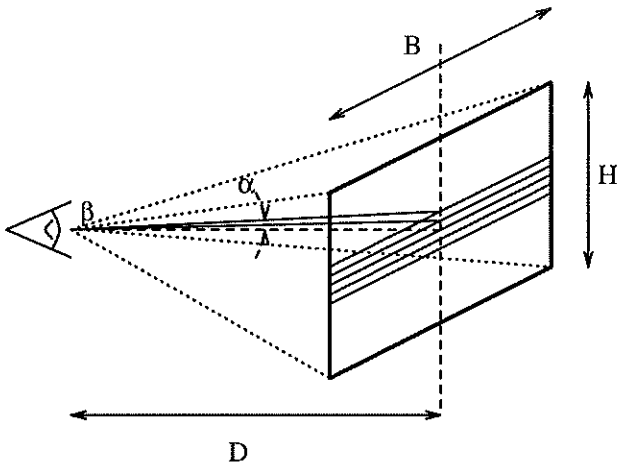


Figure 4.2: *Angle and distance of sight.*

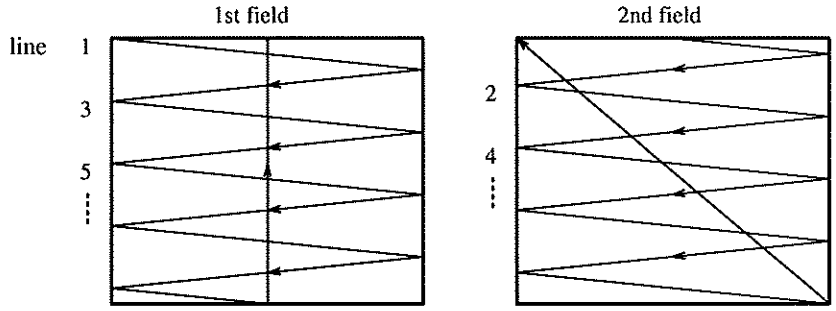


Figure 4.3: *Division of the raster into fields.*

the following (rather rough) approximation for the minimum number of lines is derived:

$$L = \frac{2500}{D/H}$$

This means a number of 500 visible lines. In accordance with CCIR [1] the complete raster area has been divided into 625 lines, 575 of which are in the visible area due to the vertical flyback of the beam during which the remaining lines are lost (we'll come back to this soon).

To reproduce a continuous rapid movement, a certain minimum frequency is required, so that no annoying discontinuities occur. For cinema ($35mm$ film), this is 24 frames per second. Considering the linkage to the AC supply frequency, a picture repetition of $25Hz$ for $50Hz$ AC supply has been selected. ($30Hz$ for $60Hz$ AC supply in America). This is however not sufficient for a flickerfree reproduction of the picture (see chapter 3). This phenomenon can be significantly reduced by interlaced scanning. The lines of the complete raster are divided into two fields, which are interlaced and transmitted consecutively [2].

[1] Commité Consultative International de Radiation

[2] Interlaced presentation originated from the field of motion picture projection, where one used to display

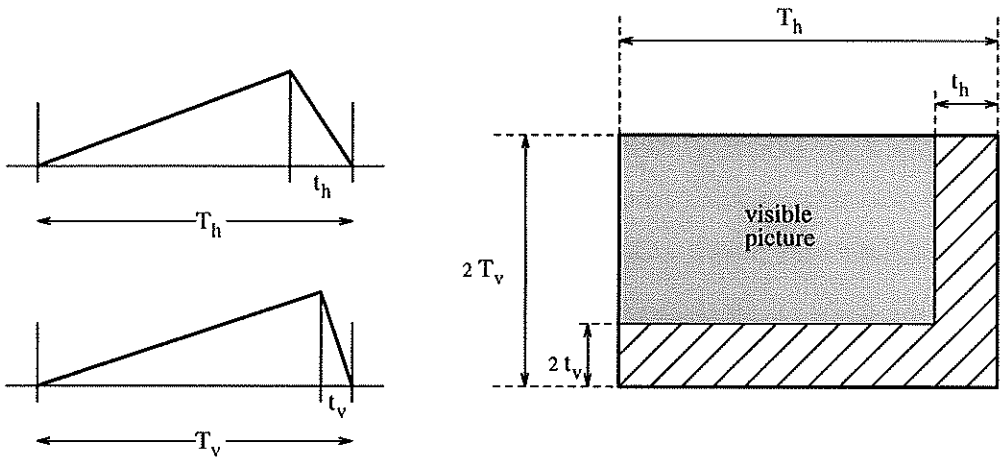


Figure 4.4: *Flyback intervals and visible raster area.*

Each field contains $\frac{L}{2}$ lines and is swept within the interval T_v. The fields should be accurately interlaced, in order to avoid pairing of lines. When using an odd line number, such as 625, the transition from the first to the second field takes place after the first half of the last line in the first field. Illustrations for such fields are provided by fig. 4.3. The flyback to the start position of the next field can be accomplished by providing periodic and fixed flyback translations of the beam: vertical flyback from the 1st to the 2nd field and vertical + horizontal flyback from the 2nd to the 1st. Therefore, no special auxiliary signal is required for switching between odd and even fields. Each second 50 fields of $312\frac{1}{2}$ lines are transmitted. Hence, the vertical frequency or field-repetition is $f_v = 50Hz$. The resulting line or horizontal frequency is

$$f_h = 1/T_h = 25 \times 625 = 50 \times 312\frac{1}{2} = 15625Hz$$

In the US the RETMA [3] standard of 525 line frames at 30 Hz is used, with 2 interlaced 262.5 line fields at 60 Hz.

At the end of a line the scanning beam is returned to the left and after sweeping a field to the top of the raster. During flyback the writing beam is blanked, which requires a fraction of the horizontal deflection time T_h and vertical deflection time T_v, respectively. In accordance with CCIR the flyback times are:

$$t_h = 0.18T_h = 11.52\mu s$$

$$t_v = 0.08T_v = 1.6ms$$

Thus, for transmitting only the line interval $T_h(1-0.18) = 52.48\mu s$, and $L(1-0.08) = 575$ lines can be used, and the visible area is reduced. Similarly, the number of visible lines for the RETMA standard (US) is reduced to 484 lines.

each shot two or three times.

[3]Radio Electronics Television Manufacturing Association

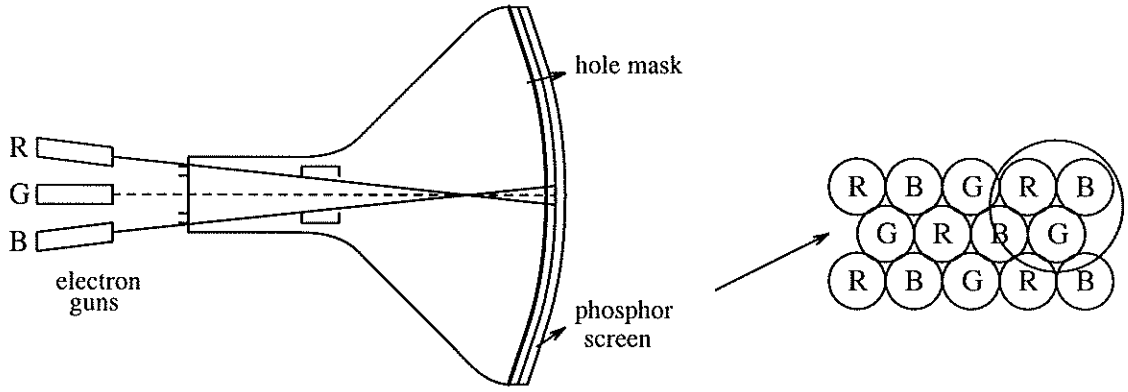


Figure 4.5: *Colour picture tube.*

Both standards have a 4 : 3 rectangular format. This aspect ratio is outside the optimal range given at the introduction of the section. Thus the number of (visible) picture elements per line is

$$\frac{4}{3} \times 625(1 - 0.08) = 767$$

and the total number of pixels is $767 \times 625(1 - 0.08) = 440833$. The time available for scanning one element is

$$T_{PE} = \frac{52.48\mu s \times 575\ lines}{440833} = 0.0684\mu s$$

and thus the highest picture signal frequency (i.e. one cycle per two subsequent pixels) would be

$$T_P = 2T_{PE} = 0.137\mu s \quad f_P = 7.3MHz$$

However, considering the fact that the beam diameter is finite, the resolution is reduced, which is expressed by the Kell factor $K \sim \frac{2}{3}$ (in a rather handwaving way for that matter). According to CCIR, the resulting bandwidth is

$$B = 5MHz$$

4.2.3 Colour CRTs

The colour picture tube has three beam producing systems. Each controls the intensity of one colour: red, green, or blue. Indeed, the colour CRT has three kinds of phosphors, one of each of these colours. There are good reasons why specifically red, green, and blue are used. As is explained in chapter 13 such choice guarantees that the widest possible gamut of colours can be produced.

In the classical tube these three systems are arranged at angles of 120° to one another. The intensity of the emitted electron beams is controlled by the EBU R_r, G_r, B_r tristimulus signals that are applied (see chapter 13). The colour resolution is determined by the size

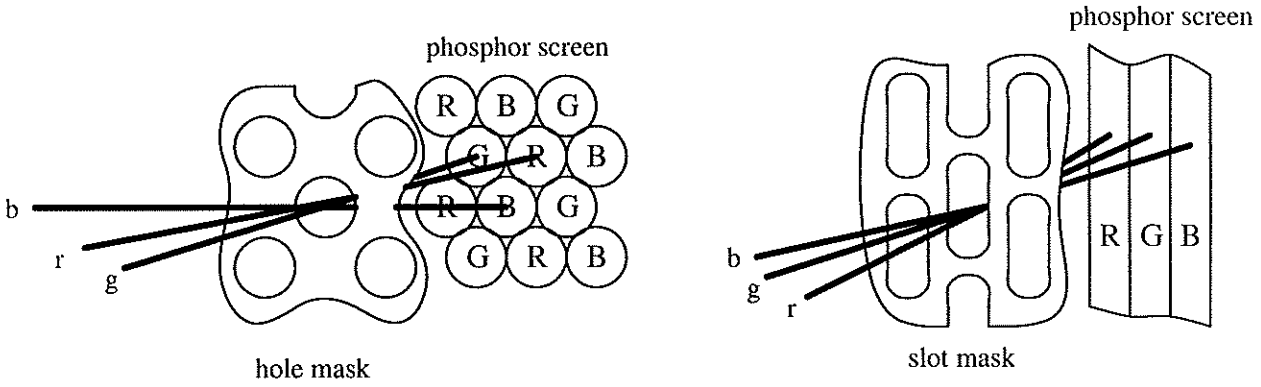


Figure 4.6: *Left: hole mask in delta picture tube, right: slot mask.*

of the individual pixel elements. Each pixel element is represented on the screen by three luminiscent dots, which are red, green and blue. To obtain a clear assignment to the corresponding dots, a hole mask is placed about $15mm$ from the phosphor screen (so-called delta picture tube). More recent systems (e.g. computer workstations) have a slot mask. The luminiscent areas on the screen are then in the form of stripes. Both configurations are shown in fig. 4.6.
With slot masks the three electron guns can be arranged in one plane. This configuration has the advantage that it leads to a high colour purity, i.e. excellent convergence is achieved and the electron beams strike the right luminiscent stripes precisely.

4.3 I - Optics for image formation

In chapter ?? the optical effects of reflection, refraction, scattering, and absorption have been discussed. These effects are used to manipulate light as to guide it in appropriate ways to illuminate the scene or to form an image on a light detector. Lenses, mirrors, and filters are the most important tools. Here we focus on the image formation aspect.

Wanted and unwanted effects go hand in hand and optical design is a balancing act between considerations such as the geometrical precision of the resulting image and other aspects like different wavelengths behaving differently (dispersion!), the amount of light that in the end strikes the light sensitive area of the detector, and ... cost.

4.3.1 Lenses

The camera obscura If the process of image formation is manageable at all, it is because light propagates along simple straight paths. This observation is the basis of the *camera obscura* [4]. In this simple predecessor of the camera, an image of the outside world

[4]The principles of the camera obscura, viz. image formation by means of a small hole, was already known to the Greek natural philosopher Aristotle (384–322 BCE). The first known (extensive) technical description of the pinhole camera was given ca. 1500 by the Italian artist and engineer Leonardo de Vinci

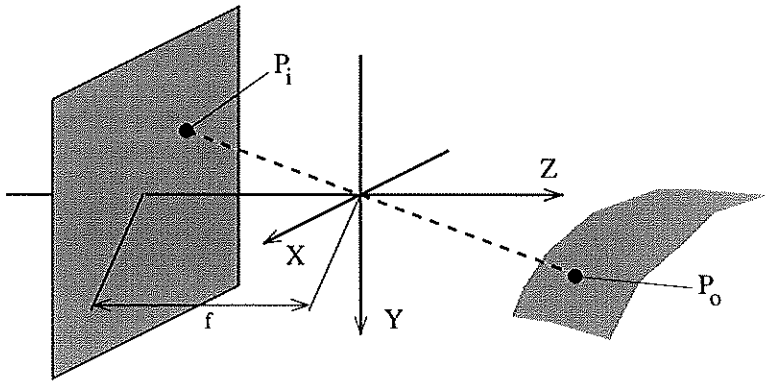


Figure 4.7: *The pinhole camera model.*

is formed on a plane wall, the image plane, that is enclosed in a completely opaque and reflectionless room or box. Light is only piercing in through a minute hole opposed to the image plane. The light coming from (reflected by) a point $P_o(X_o, Y_o, Z_o)$ on some object in the direction of the hole $O(0,0,0)$ reaches the image plane in a point $P_i(X_i, Y_i, -f)$ as indicated in fig. 4.7. This simple setup corresponds to the *pinhole camera model* and the image formation geometry is referred to as *perspective projection.* As long as the diameter of the hole is sufficiently large compared to the wavelength of the light, the influence of diffraction will be negligible, and the light may be assumed to follow a straight path. From similar triangles we find

$$\frac{X_i}{X_o} = \frac{Y_i}{Y_o} = -\frac{f}{Z_o} = -m.$$

m is called the *linear magnification.*

The lab experiment of fig. 4.8 illustrates the camera obscura. The components in this setup were from left to right: a point light source, a mask with an 'F'-shaped cut-out, a diaphragm that is almost closed to form a small aperture, and the image (very faint). The light source was the only illumination used and the surrounding walls were covered with back cloth to further suppress ambient light. Nevertheless, the intensity of the image can be seen to be low.

The thin-lens equation Of course, the "hole" in our camera (obscura) cannot be made very small, otherwise insufficient light (photons) would reach the image plane (and additional problems with diffraction would pop up as well). This could be seen in fig. 4.8. On the other hand, a large hole would create unsharp images, because light from different points in the scene would reach the same image point. The solution is the use of a lens, as in fig. 4.9.

(1452–1519); and in 1550 Girolamo Gardano described the use of a pinhole camera with a lens. In those times the camera obscura was mainly used for amusement and as a drawing tool. The oldest known use for scientific purposes dates from 1544. In that year a pinhole camera was used at the Catholic University of Leuven to observe the eclips of the sun of January 24th, 1544. A report of this experiment is given in the book *"De radio astronomico et geometrico liber"* by the Dutch scientist Gemma Frisius.



(a)

(b)

Figure 4.8: *Implementation of a camera obscura. (a) shows the overall setup with from left to right a point light source, a mask with an 'F'-shaped cut-out, a diaphragm that is almost closed to form a small aperture, and the image (very faint). (b) shows the mask, aperture and image in more detail.*

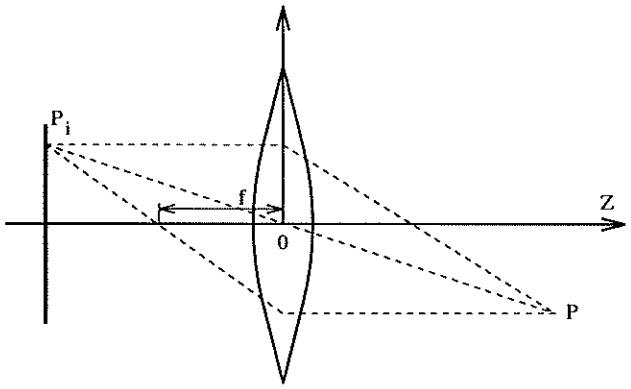


Figure 4.9: *Camera model with lens.*

The figure shows a section through the *optical axis*, i.e. the Z-axis, that passes through the center of the lens. The lens surfaces are assumed spherical on both sides. All light coming from a point P and falling on the lens will now contribute in the brightness formed at P_i. The geometry of the projection remains the same, as can be seen from the ray through the optical center of the lens, which is still unbroken. The price we pay is that only points at a certain distance will be sharply imaged. For thin lenses this distances follows from Gauss' formula:

$$\frac{1}{Z_o} - \frac{1}{Z_i} = \frac{1}{f}, \tag{4.1}$$

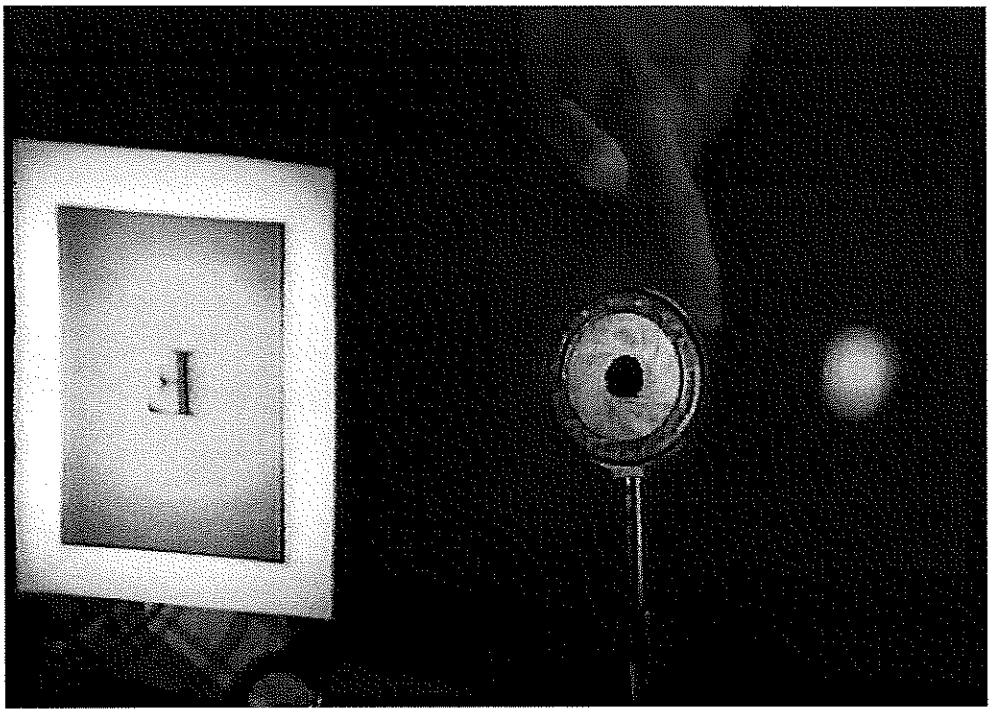
specifying that a sharp image of a point with Z-coordinate Z_o will be formed at (negative) coordinate Z_i, i.e. distance $|Z_i|$ from the optical center of the lens. The constant f is called the *focal length*. If the image is placed at a distance f from the optical lens center, points at infinity (and at large distances in practice) are focused.

The lab experiment of fig. 4.8 is continued in fig. 4.10. As part (a) shows, widening the aperture is no solution, as the image is completely blurred. A lens – added in part (b) – keeps the image sharp while allowing more light to pass through the widened aperture.

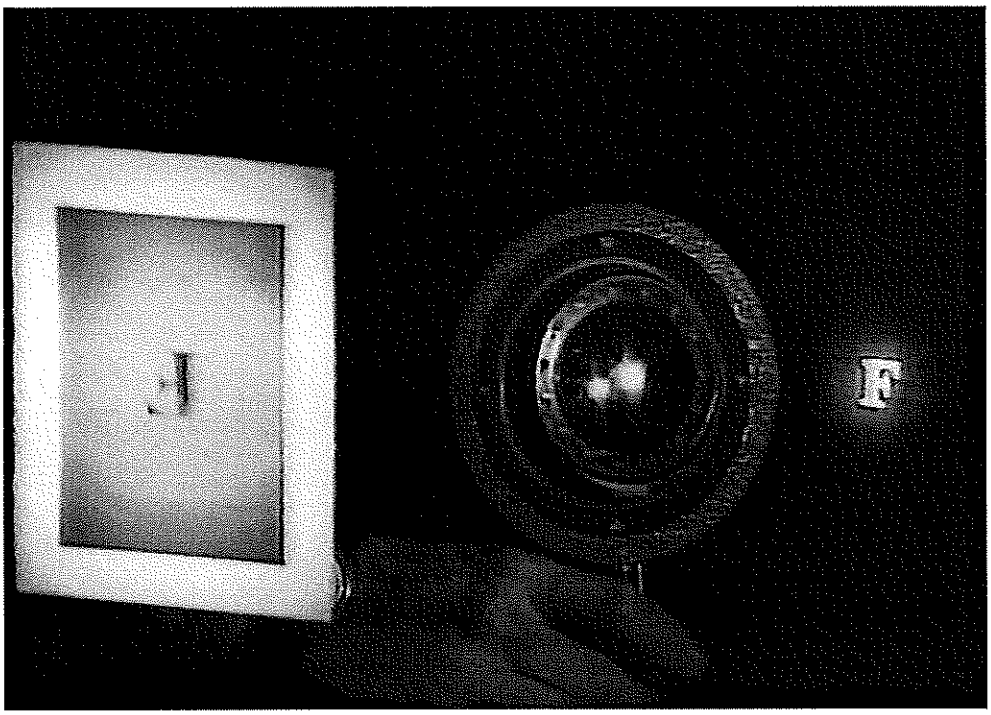
It is important to keep in mind that the derivation of eq. (4.1) the following assumptions are made:

1. the lens surfaces are spherical, as mentioned,
2. incoming light rays make a small angle with the optical axis,
3. the lens thickness is small compared to its radii of curvature, and
4. the refractive index is the same for the media on both sides of the lens.

The equation owes it name to the third restriction. The second condition is automatically satisfied if the distance of objects to the lens are large compared to its diameter. The fourth condition is satisfied if the lens is placed in a homogeneous medium (e.g. air).



(a)



(b)

Figure 4.10: *Implementation of a camera obscura cont'd. (a) shows the effect of widening the aperture. More light passes through but the image is completely blurred. (b) shows how the addition of a lens yields a sharp image that is much brighter than the original one.*

The lensmaker's formula If the refractive index of the surrounding medium is n_s and the refractive index of the lens material is n_l, then it can be shown that

$$\frac{1}{f} = \left(\frac{n_l}{n_s} - 1\right) \left(\frac{1}{r} + \frac{1}{r'}\right) ,$$

where r and r' are the radii of curvature of the two spherical sides of the lens. This expression is usually referred to as the *lensmaker's formula.* In order to avoid confusion, it should be mentioned that in contrast to the definition used here, r' (the radius of the lens surface closest to the image plane) is most often defined to be negative, resulting in a sign change in the above equation. Therefore, it is important to keep in mind that in the forementioned formula r' was taken to be positive.

The reader can check that the magnification for the lens can also be written as

$$m = \frac{f}{Z_o - f} .$$

Note that if f is made larger, then the numerator increases while the denominator decreases. Hence, m goes up with f. For remote objects (Z_o very large), this increase is almost linear, whereas for objects that are near, the increase of m is stronger than linear. On the other hand, if an inspection problem requires a certain magnification m in order to measure certain details with sufficient precision and if the object has to be placed at a distance Z_p, then the corresponding focal length is found to be

$$f = Z_o/(1 + 1/m) .$$

The f-stop or f-number The lens was introduced in order to let more light reach the image plane and at the same time keep the image sharp. How much light is actually allowed to fall onto the light detecting area of the camera should be made dependent on the intensity of scene illumination. Over- or underexposure should be avoided, i.e. an appropriate amount of incoming light flux is gathered during the light formation time. It will in general not be necessary or desirable to collect all the light falling on the lens. Therefore, cameras are equipped with an *aperture stop* or *diaphragm* that limits the amount of light reaching the sensitive area. A measure of how much light is *blocked* is the so-called *f-stop* or *f-number.* It is defined as

$$f - stop = f/d ,$$

the ratio of the focal length and the diameter of the aperture. The amount of incoming radiation decreases with the square of the $f - stop$. The values used and indicated on lenses are 2.8, 4, 5.6, 8, 11, 16, and 22. These are rounded numbers for changes with a factor of $\sqrt{2}$ between two subsequent values. They are traditionally denoted by $f/2.8$, $f/4$, etc.

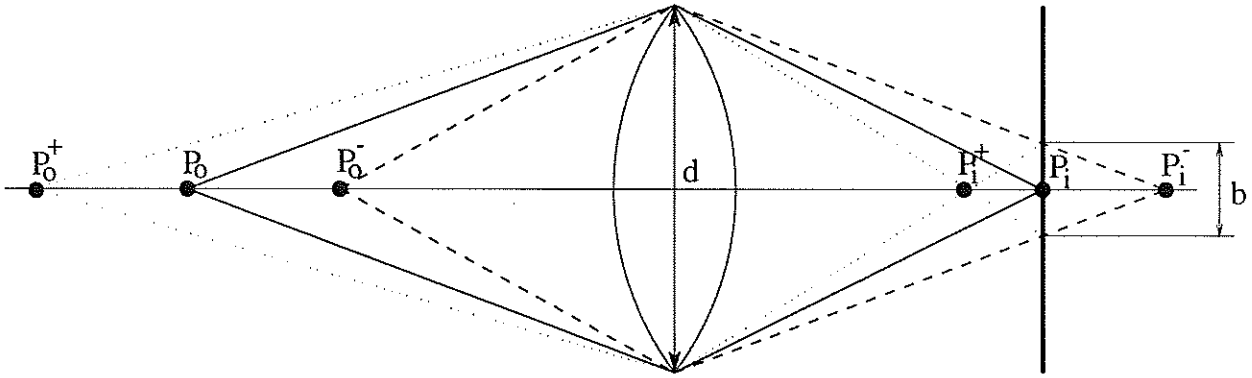


Figure 4.11: *For a certain range of positions Z_o a point is imaged as a spot of diameter smaller than b.*

The depth-of-field Given f and Z_i, points at other distances than that specified in formula (4.1) will become unsharp spots in the image. This degradation can be modelled as a convolution with a blurring filter (see chapter ??). In practice, a whole range of distances will be projected with acceptable sharpness, due to the finite resolution of image formation and human perception. Indeed, suppose an image is called sharp or "in focus" as long as a point projects to a spot with a diameter below some (small) value b. For the value of b one can choose the resolution of the sensing device. Suppose the image plane is at Z_i and the point position for a sharp projection following eq. (4.1) is at Z_o, as shown in fig. 4.11. How far can the point be moved out of this position before the focused image turns into a circular spot of diameter b, the blurred projection of the displaced point? For instance, shifting the point farther from the lens to the position of Z_o^+ will no longer yield a focused image in the image plane containing P_i, but at a point P_i^+. Similarly, shifting the point to the new position Z_o^- will yield a focused image at Z_i^- instead. In the figure the positions P_o^+ and P_o^- have been chosen as to generate a circular image of diameter b on the image plane, i.e. the correspond to the boundaries of the interval where we would consider the image to be in focus, i.e. they determine the *depth-of-field*. To find the extent of the depth-of-field, we have to calculate $\Delta Z_o^+ = Z_o^+ - Z_o$ and $\Delta Z_o^- = Z_o - Z_o^-$, where Z_o^+, Z_o, and Z_o^- are the (positive) Z-coordinates of the points P_o^+, P_o, and P_o^- as drawn in fig. (4.11).

First, we define $\Delta Z_i^+ = |Z_i^+| - |Z_i|$ and $\Delta Z_i^- = |Z_i| - |Z_i^-|$, with Z_i^+, Z_i, and Z_i^- the (negative!) Z-coordinates of the points P_i^+, P_i, and P_i^- as drawn in fig. (4.11). The principle of similar triangles yields

$$\frac{\Delta Z_i^-}{b} = \frac{|Z_i| + \Delta Z_i^-}{d} \ ,$$

or,

$$\Delta Z_i^- = \frac{b}{d-b}|Z_i| \ , \tag{4.2}$$

a relation that will be used in a moment. Considering that

$$\frac{1}{Z_o^-} + \frac{1}{|Z_i^-|} = \frac{1}{f} \ ,$$

it is rather straightforward now to make the following deductions:

$$\begin{aligned} Z_o^- &= f\frac{|Z_i^-|}{|Z_i^-|-f} \\ &= f\frac{(|Z_i|+\Delta Z_i^-)}{(|Z_i|+\Delta Z_i^-)-f} \\ &= f\frac{|Z_i|+b|Z_i|/(d-b)}{Z_i|+b|Z_i|/(d-b)-f} \\ &= f\frac{1+b/(d-b)}{1+b/(d-b)-(f-Z_o)/Z_o} \\ &= f\frac{dZ_o}{bZ_o+f(d-b)} \end{aligned}$$

Then,

$$\Delta Z_o^- = Z_o - Z_o^- = \frac{Z_o(Z_o-f)}{Z_o+fd/b-f} \ .$$

A similar expression can be obtained for ΔZ_o^+. Lowering d (e.g. by narrowing the diaphragm) is seen to enlarge the depth-of-field. Of course, one has to strike a balance between a higher depth-of-field and less light reaching the image plane. The depth-of-field is also larger for points that are farther from the camera (larger Z_o).

In summary, introducing a lens to some extent solves the problem of insufficient light reaching the light detecting area of the camera. The price we pay is the loss of overall sharpness, i.e. points at a distance outside some range are no longer in focus.

Aberrations

In the above model three important assumptions were made that are not perfectly matched by real lenses or lens systems:

1. all the rays emerging from a point in space are focused onto the same image point,
2. the image points all fall in a single image plane,
3. magnification is constant throughout the image plane.

The deviations from this ideal are called *aberrations.*

A first type of aberrations are geometrical in nature. These remain small for *paraxial* rays, i.e. rays close to the optical axis, since the paraxial approximation underlies most of the formulas of geometrical optics. Incoming rays at the outer portions of the lens do not fulfill the assumptions so well and deviations from the simplified model become visible. *Spherical aberration* is the phenomenon of incident rays parallel to the optical axis not

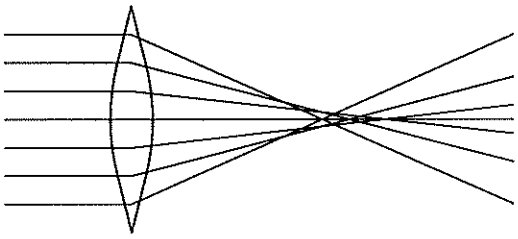


Figure 4.12: *Spherical aberration.*

converging at a single point. As seen from fig. 4.12 the outer portions of the lens yield smaller focal lengths.

Astigmatism is a similar effect for rays that are inclined with respect to the optical axis. Suppose a cylindrical beam of such rays falls upon the lens with its central axis in a *meridional plane*, i.e. a plane containing the optical axis. After being refracted by the lens, the beam will be no longer cylindrical. Focusing is realized in the *meridional plane* and the *sagittal plane* (the plane containing the beam axis and perpendicular to the meridional plane), but at different distances from the lens.
The magnification of the system is also different for different angles of inclination to the optical axis. As a result, shapes will be deformed, an aberration referred to as *distortion*. Distortion can be assessed by taking the image of a regular pattern.
Spherical aberration depicts a point as a circle. *Coma* is the effect that a point not lying on the optical axis is depicted rather as a complicated, asymmetrical comet-shaped blob. Geometrical distortion can be considerably weakened by using lens systems rather than single lenses, and by carefully selecting their characteristics and the materials they are produced from.

The second type of aberrations are *chromatic*. They find their origin in the dependence of the refractive index on wavelength. As a consequence, rays of different wavelengths will be focused in different planes. Chromatic aberration cannot be removed completely. Usually the focal lengths for two wavelengths are made equal. Again this is achieved by using appropriate systems of lenses. Sometimes *achromatization* is attained for more wavelengths. The choice of these wavelengths will depend on the application, i.e. the part of the spectrum that is most relevant.

Practical lenses Single lenses are rather rare in practical systems. Usually, multiple lenses are combined to deal with aberrations. Systems with 10 lenses – which may be of materials with different refractive indices – are not exceptional. Lenses also will be covered with anti-reflection coatings. If every single lense reflects only a few percent of the incoming light, then the system will result in a substantial reduction. Coatings can e.g. reduce reflection by glass lenses from around 4% to around 1.5%. Usually the major concern are not the reflection losses themselves but rather the stray light they cause. Calculations for lens systems are more complicated, and methods based on matrix calculations have been devised to track paraxial rays (i.e. rays close to and with small inclination angles

with respect to the optical axis). Nowadays several computer packages are available that assist in geometrical design.

If large apertures are required, e.g. for directing light from a light source, then lenses can get bulky and heavy. The solution can be to use a *Fresnel lens.*

*** Lens materials** The typical material used for the manufacturing of lenses are optical glass, acrylic plastic, fused silica, fused quartz, calcium fluoride, germanium, zinc selenide, and sapphire.

The optical features specified in catalogues usually are the refractive index and Abbe number. These are defined on the basis of some well-defined wavelengths: 486.13 nm indicated by F, 587.56 nm indicated by d, and 656.28 nm indicated by C. The "refractive index" is defined as the refractive index n_d at 587.56 nm and the Abbe number V_d is given by

$$V_d = \frac{n_d - 1}{n_F - n_C} \ .$$

Sometimes the refractive index and Abbe number are combined into a 6 digit code, the first three specifying the first three digits after the decimal point of n_d and the last three giving the three most significant digits of V_d. For instance, the optical glass BK7(517642) has $n_d = 1.517$ and $V_d = 64.2$. For most applications, a good lens ought to have both a high refractive index and Abbe number. This allows to obtain components with a high refracting power but small curvatures. In addition some of the aberrations, and mostly the chromatic aberration, can be kept small. If cost is not to increase too much, an increase in one of these parameters typically implies a decrease in the other, however.

The most often used material for lenses are the optical glasses, although the recent development of low-priced consumer market cameras shows an increased use of plastics. Optical glasses strongly dominate lenses for the visible spectrum. They are called *crown glasses* if they have an Abbe number higher than 55, or higher than 50 if their refractive index also is higher than 1.6. Other glasses are called *flint glasses.* Crown and flint glasses are combined in compound lenses to correct chromatic aberrations.

Fused quartz and especially fused silica are useful if transmission of UV-light is required. Both materials are more expensive than optical glass, however. In the visible range these material can be preferred over optical glasses if the system has to operate in a harsh environment or under high temperatures, and the cheaper quartz is preferable. They are suitable choices for high power light sources, for instance.

Calcium fluoride is useful over an extremely large range of wavelengths, covering the UV, visible, and infrared spectra. It may replace fused silica for UV-operation. Its anomalous dispersion characteristics render it useful for the correction of chromatic aberrations. It is the cheapest material if transmission in the visible range has to be combined with the infrared up to 5 μm. Humidity and high temparatures may pose problems.

Germanium has a high transmittance for infrared radiation in the range from 1.8 μm to 12 μm. It is opaque for visible radiation. It is thus useful for both the $3 - 5$ μm and $8 - 12$ μm windows of atmospheric transmittance. It has a high refractive index and high

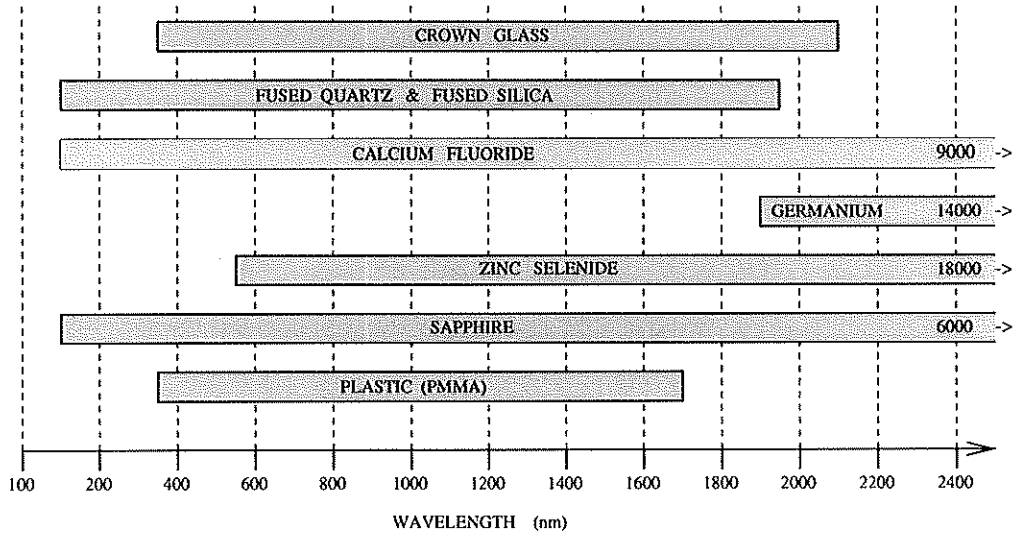


Figure 4.13: *Spectral transmission ranges for different optical material.*

refractive power can be obtained with rather thin and weakly curved lenses. A resulting disadvantage is the high level of reflection from each air/germanium surface (about 36 %!). For most applications anti-reflection coatings will be required. Germanium also poses problems if used at temperatures above $50°C$.

Zinc selenide lends itself well for transmission of infrared radiation, also at higher temperatures. With its transmission range of $0.5 - 15\ \mu m$ it is also transparent for part of the visible spectrum, a feature that simplifies the alignment of lens systems. The absorption losses are lower than with germanium and with its lower refractive index reflection at the interface with air is reduced to 17%.

Sapphire has excellent optical and mechanical properties. It transmits well from UV up to infrared ($0.18 - 5.5\ \mu m$). It also is very expensive and its use is therefore limited to particular applications, e.g. as a window shield to protect more vulnerable optics from harsh environments.

Acrylic plastic lenses (PMMA: polymethylmethacrylate) have the shortest wavelength interval where they are transparent. These materials are cheap and weigh less, and lend themselves well for the production of Fresnel lenses.

Fig. 4.13 gives an overview of the transmission windows for the different optical materials discussed.

4.3.2 * Mirrors

Not available yet.

4.3.3 * Filters

Not available yet.

4.3.4 * Optical fibers

Not available yet.

4.4 I - Cameras

4.4.1 * Classification of photosensitive devices

The light sensor (at the image plane for a focused camera) converts incoming electromagnetic energy to an equivalent response that is easily measurable. One distinguishes thermal light sensors and quantum light sensors. Thermal sensors provide a photon energy to heat conversion and don't need to concern us. Quantum sensors convert electromagnetic energy to an electric signal via the generation of charge carriers in response to incident photons.

Here we focus on quantum sensors and cameras in particular.

4.4.2 Camera types

Vacuum tube cameras and solid-state cameras are the two most widespread families. The first type operates in a scanning mode. The best known representative of this class is the *vidicon*. A photosensitive surface in the vacuum tube is excited by incoming photons. The excited carriers decreases the resistance of the photosensitive material, which is high in the dark in order to minimize the so-called *dark current*. The photosensitive surface is scanned by an electron beam that deposites charge. The spatial pattern of conductivity, a direct result of the pattern of incoming flux, is thereby translated into a temporal pattern of current. This signal is amplified. Other vacuum tube cameras differ in the photosensitive material used. Several types are available, e.g. the plumbicon and newvicon, that have each their own advantages and disadvantages.

Solid-state cameras have a separate small photosensor at each position, arranged in a regular array. These charge-coupled devices (*CCDs*) are provided with a storage potential well and can therefore collect minority carriers. These are generated by incident flux. The collected charges are shifted out one by one. No scanning is involved. There are two large families of CCDs: area CCDs and linear CCDs.
With area CCDs the photosensitive elements are arranged as a two-dimensional area. There are two basic area CCD architectures: *frame transfer* CCDs and *interline transfer* CCDs. Figure 4.14 (a) and (b) show the interline transfer CCD and frame transfer CCD, respectively. The FTCCD consists of three major blocks: the image section, the storage section, and the serial readout register. Both the image and storage section consist of vertical CCD registers that are placed side by side. The serial readout register is placed horizontally. Only the image section is light sensitive. The rest of the chip is shielded from the incident light. After the integration time the information is rapidly shifted from the image section to the storage section. Immediately after this a new integration period can be started. The charge in the storage section is shifted down, a line at a time, into

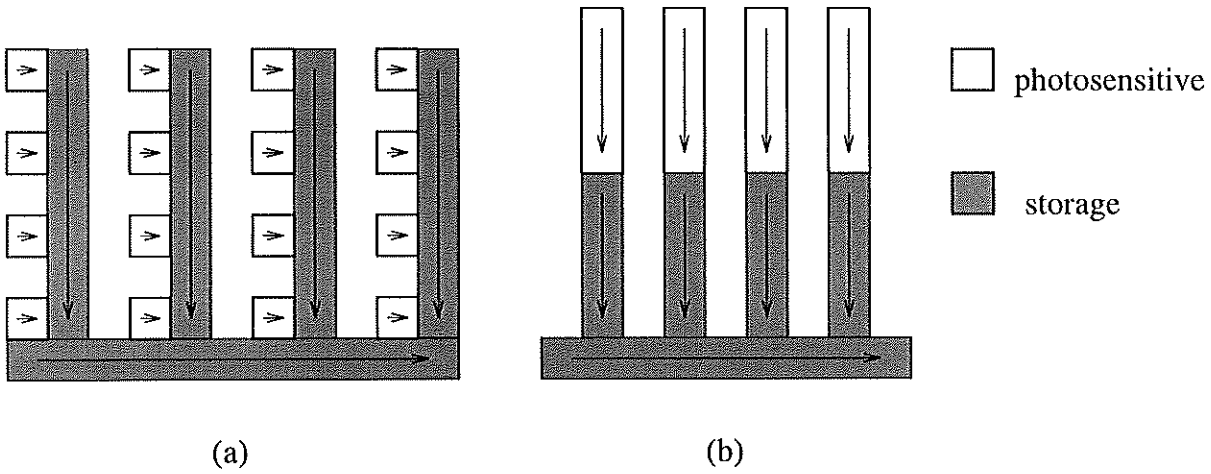


Figure 4.14: *Two types of CCD arrays: (a) interline transfer CCD, (b) frame transfer CCD.*

the horizontal readout register, which shifts it, pixel per pixel, to the output structure. The FTCCD is the highest resolution architecture. Furthermore, the entire pixel area is photosensitive such that the average quantum efficiency is high.
The ITCCD consists of an alternating structure of vertical CCD shift registers used for charge transfer and columns of photodiodes for light sensing. After the integration period all the charge packets collected on the photodiodes are transferred to the vertical CCD registers. The latter are covered by metal light shields. During the next integration period the vertical CCDs transfer their charge packets, one line at a time, to a horizontal readout CCD that transports the charges to an on-chip readout structure. The ITCCD results in lower density, but the use of photodiodes as sensing elements provides a better sensitivity for blue light.
Solid-state linear sensors consist of linear arrays of photo-sensitive elements. Using the motion of objects on a conveyor belt for the perpendicular direction, complete images can be formed. Resolution can be very high for a reasonable price (e.g. 4096 elements). As the light-sensitive elements take less place as with matrix cameras, extra logic can more easily be placed on the same chip. Some linear CCDs have their own microprocessor, and can perform e.g. (simple) edge detection.
In the future, pixel processing operations will more and more be performed on the sensor chip itself (smart sensors).

Characteristics of vidicons

Vidicons are usually operating at video signal standards (NTSC for the United States and Japan, CCIR for Europe). Unfortunately, the 4:3 aspect ratio prescribed by these standards is not really optimal for a host of image processing applications, where square images would be more appropriate. Also interlacing (odd and even lines are scanned as two different images) is not useful when a computer is to interprete the images. For

instance, a product to be inspected might have moved considerably in the time between the acquisition of two subsequent lines, resulting in a translation between odd and even lines. This makes mensuration more difficult. The standards also prescribe a constant resolution and constant frame time which will not always be appropriate for the particular application.

Returning to the physical characteristics of the vidicons *per se*, their non-linear response to light intensities can be a disadvantage, although, on the other hand, it increases the camera's dynamic range. Other weaknesses are: short tube life, drift and aging (response parameters change with time), significant image lag (influence of preceding images) and blooming (a bright spot spreads out its energy over an entire region), the risk of burn-in, and vulnerability to shocks and magnetic fields. They also weigh more, require more power and are larger than solid-state cameras. Their main advantages over the latter have been a lower price (as they were mass-produced for the video market contrary to solid-state cameras) and higher resolution. Both advantages have disappeared completely near the end of the 80's and by now the situation has been reversed. Vidicons have a spectral responsitivity which is mostly limited to ultraviolet and the visible spectrum.

Characteristics of solid-state cameras

Solid-state area array or matrix cameras contain area arrays of photo-sensitive elements. The major types are Charge Coupled Devices or CCDs and Charge Injection Devices or CIDs. CIDs involve exactly the same mechanism for detecting photons as CCDs. They differ only in the methods used for reading out the photo-induced charges.

Special image processing solid-state cameras with aspect ratios of 1 and square arrays with power of 2 dimensions (e.g. 512×512) are commercially available. Geometrical distortions – another important factor for metrological applications – are non-existent as no scanning is involved. Some elements may differ in area though, usually complete columns, and this causes systematic sensitivity differences which can be compensated for if necessary. Resolutions are comparable or higher than with vidicons (1024×1024 is readily available and 4096×4096 has been designed for aerospace applications).

Response to light is usually more linear. Their spectral sensitivity is less similar to the human eye, however. It is shifted towards the infrared. This feature can sometimes be useful, e.g. for welding robots when observing the weld puddle. Also their insensitivity to magnetic fields is important in such an environment. But red and infrared light causes more cross-talk between the integration elements than cold light and extra blurring as the refractive index of lenses depends on wavelength. Therefore, these frequencies can better be eliminated for high precision mensuration.

CCDs should be used when large intensity variations are to be expected. Specular reflections are a typical cause of such problems. There is no risk for permanent burn as in Vidicons. Integration time can potentially be adjusted to the lighting conditions and the required frame rate. Image lag is small. CID and photodiode arrays produce more noise than CCDs, but are less susceptible to blooming and dark current.

Solid-state cameras are ideal for hand-eye vision in robotics applications as they can be

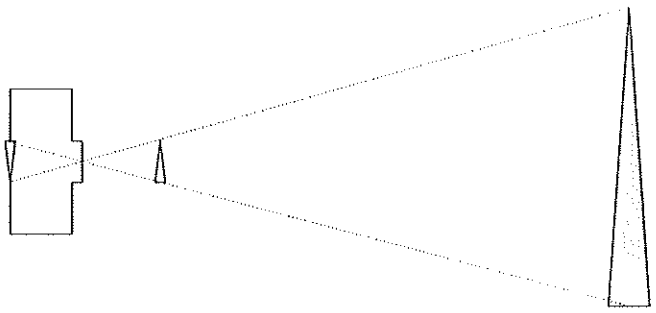


Figure 4.15: *In a* pinhole camera *an image of the scene is formed by the rays of light that are reflected by the objects in the scene and fall through the center of the lens onto the image plane. The image plane is physically located behind the center of projection (grey). The photo-positive (white) image is situated in front of the camera, where a virtual image plane can be thought to be located.*

made small, light, and robust. Some cameras have windowing features: only some parts of the image which contain the relevant information are transmitted to the vision computer, with lower frame times as a result (CID technology). This is useful for visual inspection. Sensors have even been developed that can adapt their resolution within windows upon e.g. the detection of motion.

Part on CMOS cameras not available yet

4.4.3 Models for camera projection

Perspective projection

Camera images are the typical input to a computer vision system. The pinhole camera model – introduced with the camera obscura – is our point of departure. The introduction of lenses allows us to stick to this model, at the expense of limited focus for a given focal length. An image is formed because light rays through the optical center fall onto the image plane, i.e. the plane of the light sensitive film, tube, or chip. The rays can now be thought to pass through the center of the lens, which acts as the center of projection. The situation is depicted in fig. 4.15. In the figure, two projections are shown. The one shown in grey corresponds to the projection as it will physically exist in the camera. The one shown in white is positioned at the same distance from the center of projection, but before the camera. In this *virtual* image plane, the image is upright. This virtual placement of the image plane eliminates sign reversals of the (virtual) image coordinates with respect to the original point coordinates. This placement is also justified in the sense that if one looks at a photograph or if an image is presented on a screen, one also shows the images upright as they would be obtained when the image plane were at the virtual position. This model with the virtual image plane in front of the camera is very popular in the computer vision literature. Henceforth, the image plane will always be assumed at this position.

We now derive the projection model. Consider fig. 4.16. The coordinate system

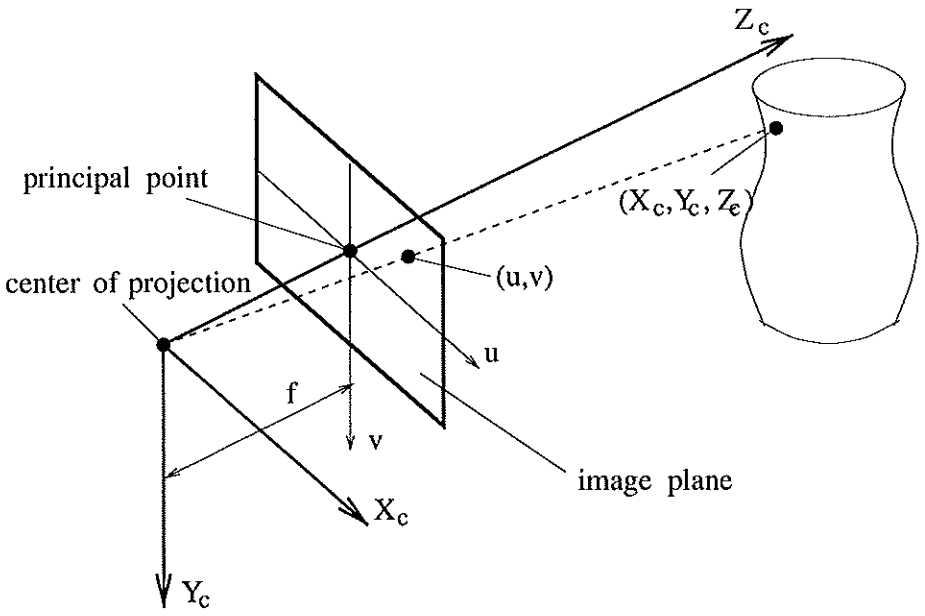


Figure 4.16: *Pinhole camera model with f the focal length. A point with coordinates X_c, Y_c, Z_c is projected at a point (u, v).*

(X_c, Y_c, Z_c) has been chosen in a specific way:

- its origin lies at the center of projection, i.e. in the center of the lens of the camera,
- the Z_c axis coincides with the optical axis of the lens or objective, and
- the plane through the center of projection and perpendicular to the optical axis is the X_cY_c-plane, with the X_c-axis parallel to the main direction, the rows say, of the image. The orientation of the Y_c is fixed by then, as it is orthogonal to both X_c and Z_c.

This coordinate frame will be referred to as the *camera coordinate frame*, hence the subscript c. Note that the u and v coordinate axes in the image plane are parallel to the X_c and Y_c axes. They intersect at the *principal point*, i.e. the point where the optical axis intersects the camera plane.

The image of a point $\mathbf{P}$ in the scene with coordinates $(X_c, Y_c, Z_c) \in \mathbb{R}^3$ is the intersection of the line through $\mathbf{P}$ and the center of projection with the image plane, i.e. the plane with equation $Z = f$). The (u, v)-coordinates of this image point in the image plane are directly found from similar triangles:

$$\begin{aligned} u &= f\frac{X}{Z} \\ v &= f\frac{Y}{Z} \end{aligned}$$

We will consider these equations as the precise model of actual projection, This model is commonly referred to as *perspective projection*. Of course, this model is an approximation

of real projection. In particular, supposing the image plane to lie a focal distance from the center of projection assumes the object to lie at a large distance compared to that focal distance.

Pseudo-orthographic projection

If we consider the above equations of projection, we see that some special cases can be distinguished. First, if Z is constant for all points of an object, then the projection model simplifies to $x = kX$ and $y = kY$ with $k = f/Z$. The effect of projection is a simple scaling of the X and Y coordinates. The object lies in a plane orthogonal to the viewing direction. Similar simplifications can be made when the range of Z values of object points is small compared to Z. In that case f/Z will remain *almost* constant and the projection model can be simplified the same way. This model for projection is referred to as *pseudo-perspective projection* or *pseudo-orthographic projection*. It can be used if the observed scene is small compared to its distance to the camera. Problems can become much easier if image formation can be approximated by this pseudo-orthographic model rather than by the full perspective model (see e.g. chapter 15). Note that in the special case where $k = 1$, the projection is plainly *orthographic*. The pseudo-orthographic model amounts to parallel projection, with a scaling added to mimic the effect that the image of objects shrink with distance.

The difference between pseudo-orthographic and perspective projection is shown in fig. 4.17, based on 3D models. Notice how the edges of the square (top images) remain parallel under orthographic projection, whereas they are no longer parallel under perspective projection. The face (second row) shows the difference between the image acquisition conditions. The orthographic image on the left corresponds to the kind of image one would obtain with a camera zooming in from a relatively large distance. The image on the right shows appreciable perspective skew, as it simulates an image taken by a camera that is very close to the subject. In this case the orthographic projection might seem the more natural one. However, in this case the perspective view was taken from a rather extreme viewpoint. In general, perspective projection is a far better model of image formation than (pseudo-)orthographic projection, see e.g. the examples in the last row.

Projection matrices

The perspective projection model we have derived still isn't a very practical tool. It is incomplete in at least two ways:

1. the coordinates of points will often be known in terms of some *world coordinate frame*, not the camera coordinate frame; we need to transform coordinates first to the latter, and

2. in the current model, the coordinates of points in the image plane are derived from camera frame coordinates; we need to express these coordinates in terms of the row and column coordinates (pixels) that we get out of the camera.

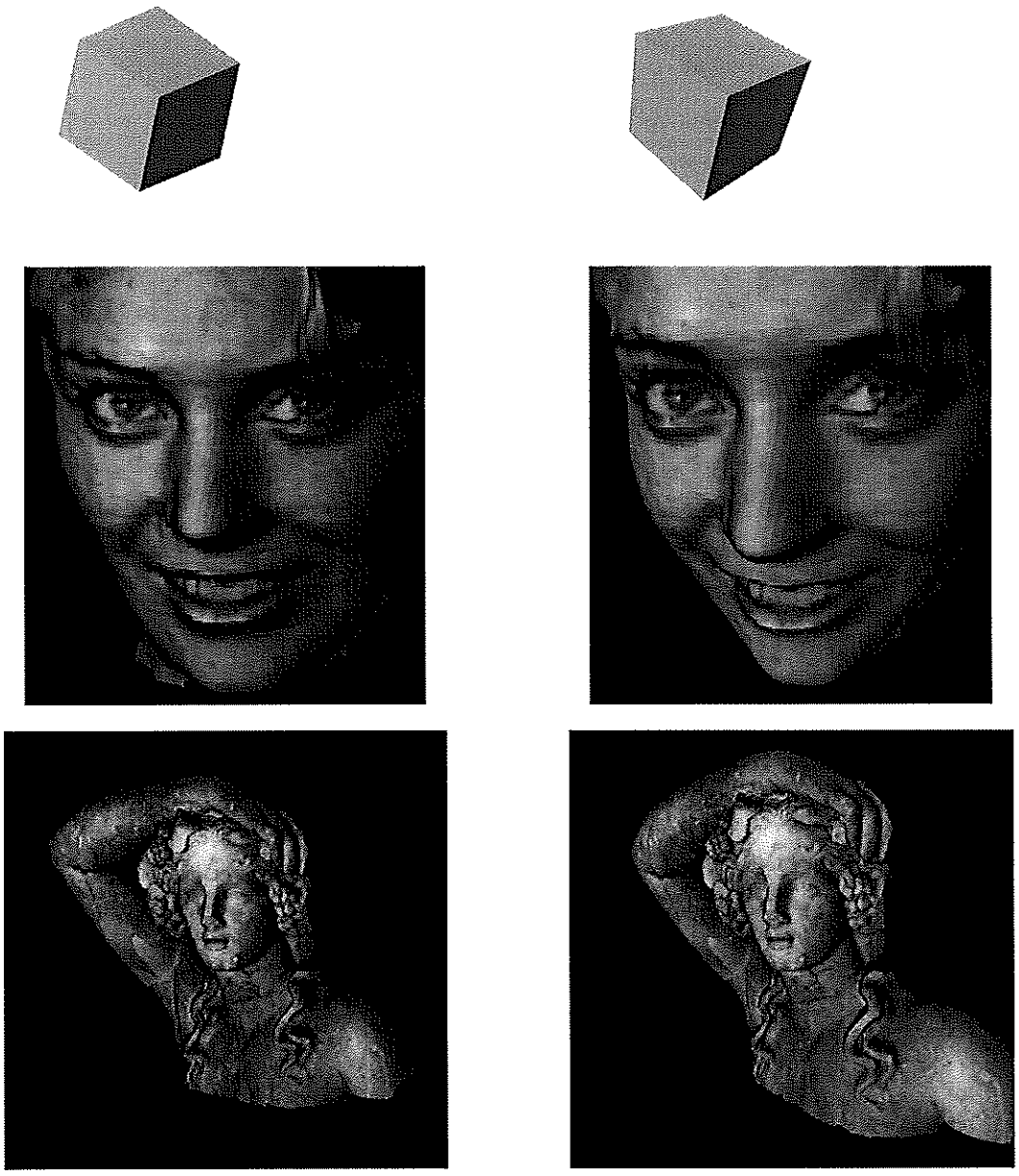


Figure 4.17: *Images generated with pseudo-orthographic projection (left) and perspective projection (right).*

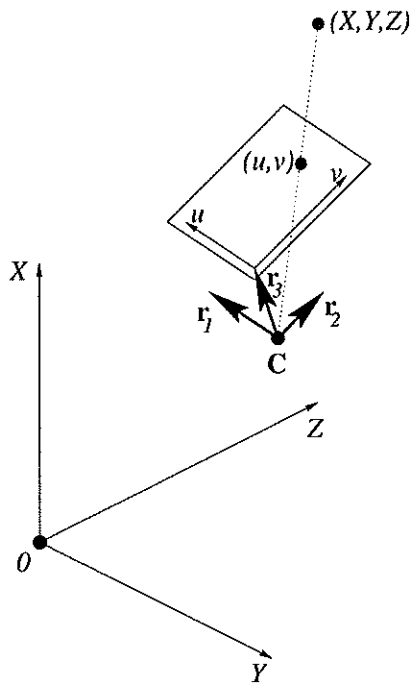


Figure 4.18: *The position and the orientation of the camera in the scene is given by a position vector* $\mathbf{C}$ *and a* 3×3*-rotation matrix* $\mathbf{R} = (\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3)$. *The image* (u, v) *of a scene point* (X, Y, Z) *is then given by formula (4.3).*

These are the aspects we deal with next. On top of that, additional refinements may be necessary to deal with geometric distortions (e.g. radial distortions) and more complicated models for the lens and the corresponding image formation process. These issues are not considered.

Suppose the position of scene points are described with respect to a world coordinate frame and we know the position and the orientation of the camera with respect to this frame. This position and orientation of the camera can be described by a point C, indicating its center of projection, and a 3×3-rotation matrix $R \in SO(3)$ indicating the orientation of the camera coordinate frame with respect to the world frame. More precisely, if $\mathbf{r}_i$ denotes the ith column of the matrix R, then $\mathbf{r}_1$ is the unit vector in the direction of the X_c-axis in terms of world coordinates, $\mathbf{r}_2$ is the unit vector in the direction of the Y_c-axis, and $\mathbf{r}_3$ is the unit vector in the direction of the Z_c-axis of the camera coordinate frame (see fig. 4.18). The coordinates of a scene point P with respect to the camera frame then are found by taking the inner products of the relative position vector $\mathbf{P} - \mathbf{C}$ with the unit vectors $\mathbf{r}_i$:

$$\big(\langle \mathbf{r}_1 , \mathbf{P} - \mathbf{C} \rangle , \langle \mathbf{r}_2 , \mathbf{P} - \mathbf{C} \rangle , \langle \mathbf{r}_3 , \mathbf{P} - \mathbf{C} \rangle \big) \ .$$

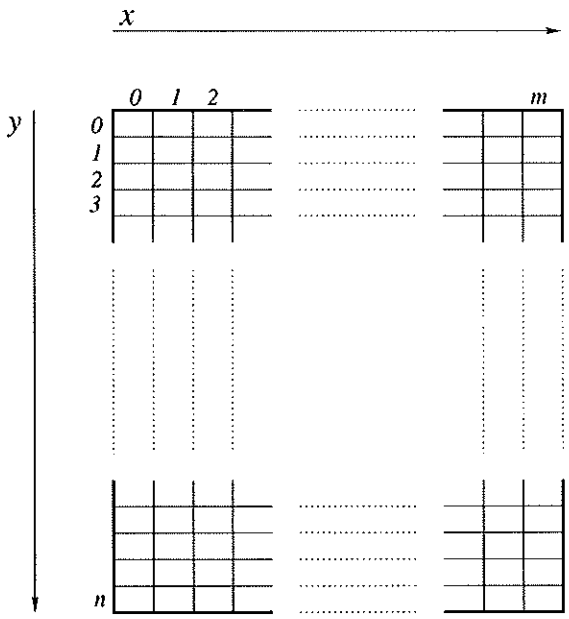


Figure 4.19: *An* $m \times n$-digital image *is composed of* mn *very small picture elements or* pixels. *It is stored in the memory of a computer as an* $m \times n$*-matrix.*

The corresponding image point then has (u, v)-coordinates

$$u = f\frac{\langle \mathbf{r}_1 , \mathbf{P} - \mathbf{C} \rangle}{\langle \mathbf{r}_3 , \mathbf{P} - \mathbf{C} \rangle} \qquad \text{and} \qquad v = f\frac{\langle \mathbf{r}_2 , \mathbf{P} - \mathbf{C} \rangle}{\langle \mathbf{r}_3 , \mathbf{P} - \mathbf{C} \rangle} \ .$$

If $\mathbf{P}$ has coordinates (X, Y, Z) and $\mathbf{C}$ has coordinates (C_1, C_2, C_3) with respect to the world frame, and if r_{ij} denotes the (i, j)th entry of the rotation matrix R, then

$$\begin{aligned} u &= f\frac{r_{11}(X - C_1) + r_{12}(Y - C_2) + r_{13}(Z - C_3)}{r_{31}(X - C_1) + r_{32}(Y - C_2) + r_{33}(Z - C_3)} \\ \text{and} \quad v &= f\frac{r_{21}(X - C_1) + r_{22}(Y - C_2) + r_{23}(Z - C_3)}{r_{31}(X - C_1) + r_{32}(Y - C_2) + r_{33}(Z - C_3)} \ . \end{aligned} \tag{4.3}$$

When working with digital images it is more natural to indicate the position of an image point in so-called *pixel coordinates*. As discussed further in chapter 5, images will have to be sampled, in order to feed them into a computer. After sampling, the image is composed of a large number of tiny, discrete *picture elements*, or *pixels* for short. This situation is depicted in fig. 4.19. A digital image thus is a 2-dimensional array of pixels, say m columns and n rows, and with each pixel is associated an intensity or 'grey level', following a process of quantization (also see chapter 5). Therefore it is preferable to indicate the position of an image point by means of its pixel coordinates (i.e. by its column and row number) instead of the (u, v)-coordinates defined above. The transition from the geometrical (u, v)-coordinates to the actual pixel coordinates, which will be referred to as (x, y)-coordinates,

is modeled by an affine transformation of the form:

$$\begin{cases} x &= k_x\, u &+s\, v &+x_0 \\ y &= &k_y\, v &+y_0 \end{cases} \tag{4.4}$$

Here

- (x_0, y_0) are the pixel coordinates of the principal point.
- k_x is the number of pixels per unit length in the horizontal direction (and thus gives 1/width of a pixel).
- k_y is the number of pixels by unit length in the vertical direction (and thus implicitly describes the 1/height of a pixel).
- s indicates how strong the shape of the pixels deviates from being rectangular. This parameter is usually referred to as the *skewness* of the pixels. Observe that $s = 0$ corresponds to rectangular pixels. This is the normal case.

Of course, in reality the pixel coordinates are discrete and small patches of the image plane will be lumped together into single pixels. These pixels have integer coordinates (column and row numbers). This discretisation is often not taken into account while performing calculations The numbers k_x and k_y depend on the focal length and the zooming distance of the lens or objective. The ratio k_y/k_x is called the *aspect ratio* of the pixels. Deviations from 1 indicate that the pixels are not square.

The numbers k_x, k_y, s, x_0 and y_0 are referred to as the *internal camera parameters.* When they are known, one says that the camera is *internally calibrated.* In practice this means that one can convert the observed (x, y)-image coordinates to metric (u, v) coordinates, and vice versa. From a mathematical point of view, one may in that case assume without loss of generality that $k_x = k_y = 1$ and $s = x_0 = y_0 = 0$. On the other hand, the vector $\mathbf{C}$ and the matrix $R \in SO(3)$ that describe the position and the orientation of the camera with respect to the world frame, are called the *external camera parameters.* If $\mathbf{C}$ and R are known, then the camera is said to be *externally calibrated.* In practice this means that one can convert the scene coordinates into camera-centered coordinates, and vice versa. From a mathematical point of view, one may in that case assume without loss of generality, that $\mathbf{C} = (0, 0, 0)$ and $R = I_3$, the 3×3-identity matrix. If the internal as well as the external parameters of the camera are known, one says that the camera is *fully calibrated*, or *calibrated* for short.

Perspective projection being a projective mapping, the image formation process in a pinhole camera (as described above) can be represented linearly by using homogeneous coordinates. Since the points considered thus far are physical points (in contradistinction to ideal points), which, moreover are given by their Euclidean coordinates, it is preferred to let the homogenizing coordinate as much as possible be equal to 1. In mathematical terms, this means that one works in a completed affine space instead of a projective one.

Explicitly, a point $\mathbf{p}$ in the image with pixel coordinates (x, y) will be represented in *homogeneous coordinates* as the column vector

$$\mathbf{p} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} ;$$

and a point $\mathbf{P}$ in the scene with scene coordinates (X, Y, Z) is represented as the column vector

$$\mathbf{P} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} ,$$

or, if *homogeneous coordinates* are needed, as

$$\tilde{\mathbf{P}} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} .$$

Equation (4.3) can then be rewritten in matrix form as follows:

$$\rho \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} fr_{11} & fr_{12} & fr_{13} \\ fr_{21} & fr_{22} & fr_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} X - C_1 \\ Y - C_2 \\ Z - C_3 \end{pmatrix}$$

where $\rho \in \mathbb{R}$ is a non-zero scalar factor. On the other hand, eq. (4.4) can be rewritten in matrix form as

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} k_x & s & x_0 \\ 0 & k_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} .$$

Defining the *calibration matrix* of the camera as

$$K = \begin{pmatrix} k_x & s & x_0 \\ 0 & k_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} fk_x & fs & x_0 \\ 0 & fk_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

yields the following *projection equations* for a pinhole camera:

$$\boxed{\rho \mathbf{p} = K R^t (\mathbf{P} - \mathbf{C})} \qquad \text{for some non-zero } \rho \in \mathbb{R} ;$$

or, if also homogeneous coordinates are used for the scene points, then the projection equations become

$$\rho \mathbf{p} = K \left(R^t \mid -R^t \mathbf{C} \right) \tilde{\mathbf{P}} \qquad \text{for some non-zero } \rho \in \mathbb{R}. \tag{4.5}$$

From linear algebra it is well-known that every 3×3-matrix of maximal rank can be decomposed as a product of an orthogonal matrix and an upper triangular matrix by means of so-called QR-decomposition. By eventually multiplying both sides of the projection equations above by the same non-zero scalar factor, it follows that KR^t, in fact, can be any arbitrary invertible 3×3-matrix M, and that $-KR^t\mathbf{C}$ can be any arbitrary column vector $\mathbf{t} \in \mathbb{R}^3$. Hence, the projection equations can formally be written as

$$\rho \mathbf{p} = (M \mid \mathbf{t})\tilde{\mathbf{P}} \qquad \text{with} \qquad \text{rank } M = 3 \ ,$$

and where $\rho \in \mathbb{R}$ is a non-zero scalar factor. The projection matrix $(M \mid \mathbf{t})$ thus is an arbitrary 3×4-matrix of maximal rank, and the previous equation describes a (linear) mapping from projective 3-space $\mathbb{P}^3$ to projective 2-space $\mathbb{P}^2$. In fact, every 3×4-matrix of maximal rank can also be interpreted as the model of a perspective projection and hence of a perspective camera.

Remark. If the object's size is small compared to its distance to the center of projection, or when a zoom lens with large focal length is used, then parallel projection offers a good approximation for the image formation process in a real camera. In the general linear camera model (as defined above) parallel projection is described by a 3×4-matrix $(M \mid \mathbf{t})$ where M is a 3×3-matrix of rank 2: its 3rd row consists of 0s.

4.4.4 From object radiance to pixel grey level

This section together with section describes the chain of phenomena that determine the brightness that an object patch will have in the image. This process consists mainly of two aspects. The first part investigates the relation between the radiance of the object's surface patch and the irradiance at the position of its projection on the image plane. The second step relates image irradiance to the actual output signal or *brightness* or *grey level* produced by the camera.

Step 1: from object radiance to image irradiance

We assume the object is opaque and we only deal with reflection. Things like local surface orientation (relative to the camera), surface reflectance, and the intensity, spectral composition and position of the light sources all have their influences. Things can get still more complicated when we consider ambient lighting, e.g. as a consequence of light being reflected by some surface parts on others (so-called interreflections). Including all these details is not very useful if at all possible and these factors will therefore be lumped together in a value for the radiance R, which will be assumed to be known.

A further simplification is to assume that the object is at a distant location compared to the focal length f of the camera. To have a sharp image the image plane should then be placed at a focal length from the camera, as shown in fig. 4.20.

We further suppose the lens has an area A_l and the normal distance between the lens

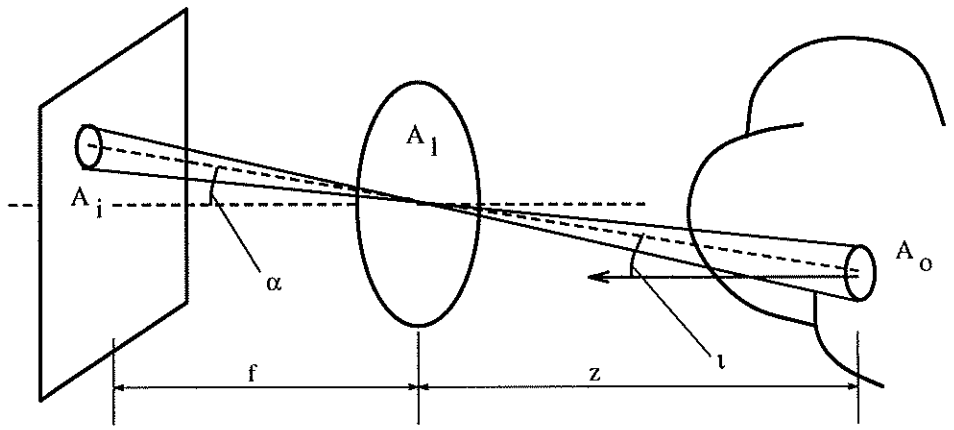


Figure 4.20: *Image formation setup.*

center and the object patch is Z. A first question to answer is what the area A_i of the patch in the image will be for a patch with area A_o on the object surface. Suppose the projection rays make an angle α with the optical axis of the camera (i.e. the axis normal to the image plane through the center of the lens). Also suppose the angle of inclination of the surface patch normal with respect to the projective ray is ι. Then the relation between the areas A_i and A_o follows from the simple scaling of the orthogonal sections through the viewing cone at both places

$$A_i \cos\alpha = \left(\frac{z}{f}\right)^2 A_o \cos\iota. \tag{4.6}$$

With an ideal lens, all light that is reflected from the patch upon the lens will hit the corresponding image patch, still assuming the camera is in perfect focus under the given circumstances. The radiance is expressed per unit of foreshortened area and per steradian solid angle in which light is to be considered. The foreshortened surface patch area, from the viewing direction, is

$$A_o \cos\iota$$

and the solid angle subtended by the lens when viewed from the surface patch is given by

$$\frac{A_l \cos\alpha}{(Z/\cos\alpha)^2}.$$

Then the incoming flux on the image patch is

$$F = R\,(A_o \cos\iota)\left(\frac{A_l \cos\alpha}{(Z/\cos\alpha)^2}\right).$$

where it is assumed that the distance from the object to the lens is much larger than the lens diameter such that ι can be treated as a constant. Consequently, the irradiance I is found as

$$I = \frac{F}{A_i} = R\frac{A_o A_l}{A_i Z^2}\cos^3\alpha\cos\iota$$

and using 4.6

$$I = R\frac{A_l}{f^2}\cos^4\alpha.$$

We conclude that the irradiance is proportional to the radiance if the viewing direction is approximately constant for all patches of the object. This is all the more true if the object is viewed along the optical axis. In that case $\cos^4\alpha$ can be approximated by 1.

Of course this result is an idealisation of reality, where one has to cope with several kinds of distortions and aberrations (see section 4.3.1). Anyway, this result suggests that if one is able to characterize the surface reflectance sufficiently to understand the relation between surface radiance and relative orientation of the incoming light from a source, then the image can be used to extract the radiance in order to invert the process and to extract the local surface normal. In general the problem with such undertaking is that the relation between the radiance and the orientation of the surface is not unique. We'll come back to this issue in chapter 11. Of course, it is not image irradiance that is given by the grey levels produced by a camera. The irradiance merely is the camera's input.

Step 2: From image irradiance to pixel grey level

In general, the grey level output of a camera for a certain position in the image is not simply proportional to the received irradiance at that point. If the irradiance is represented by I and the resulting grey level is represented by f, then the relation is of the type

$$f = gI^\gamma + d \ ,$$

with g, γ, and d constants. The factor g is the so-called *gain* of the camera. The exponent γ expresses the non-linearity of the camera response to incoming light. Only for $\gamma = 1$ is the relation a linear one. CCD cameras typically have γ values very close to 1, whereas the standard vidicon (Sb_2S_3 vidicon) has $\gamma \simeq 0.6$. For a plumbicon $\gamma \simeq 0.95$. The offset d is the *dark reference* and can be measured by taking a image with the lens cap on. It is clear that a good camera should have a small d. Several CCD cameras have photosensitive elements that are shielded off from light and thereby yield an indication of the dark reference (a consequence of the *dark current*), that among other things, depends on the temperature. The dark reference will also differ from pixel to pixel, however, and usually a number of such reference elements is provided to yield and average dark reference.

4.5 I - Scene illumination

4.5.1 * Light sources

Not available yet.

4.5.2 Overview of illumination techniques

Most vision systems work in environments that can be controlled carefully. Systems for visual inspection are a good example of this. Normally one is interested in but a number of specific features in the scene. The point is that it might be possible to highlight these without going through a painstaking process of determining preprocessing steps to do so. Simple measures such as changing the viewpoint of the camera or choosing a more appropriate type of illumination may render much of this superfluous. The following list of examples is far from exhaustive.

Backlighting

With *back-lighting* several fluorescent lamps are placed behind a transmitting diffuser plate of plastic or glass. From the viewpoint of the camera, this light source is placed behind the object. The result is a high-contrast silhouette image, which lends itself to simple, binary vision (i.e. vision with only two intensity levels, black and white). It is therefore one of the most popular techniques in industry.

Fig. 4.21 shows an example of backlighting and the difference it makes with frontlighting. The object is a board of translucent dielectric material, with metal covering. With the backlighting technique, the metal clearly stands out as black parts on a bright background. Frontlighting would make it much harder to find these parts.

Diffuse lighting

Diffuse-lighting illuminates the object uniformly from all directions. It prevents the generation of sharp shadows and large intensity variations over glossy surfaces due to specular reflections. The specular reflection from metal surfaces can easily produce a dynamic range of ten thousand to one! What diffuse lighting achieves is not so much a reduction in specular reflections, but rather the relative increase of the diffuse reflection component, yielding less variations.

Fig. 4.22 shows the improved conditions under diffuse lighting (bottom) vs. directional lighting (top), which leads to problems with large changes in intensities.

Directional lighting

Directional lighting has the opposite effect of diffuse lighting. Here it is the goal to let the specular reflection component help in the extraction of certain features. An example is the generation of specular reflection on shiny surfaces for inspection. Cracks will reflect the light in other directions than the good parts of a surface. When viewing in the direction of specular reflection for the well-finished part of the surface, the so-called *bright field* is considered. Cracks will appear as dark stripes. Viewing from other directions corresponds to investigating the so-called *dark field.* Here cracks or scratches may well appear brighter than the good surface. This is illustrated in fig. reffi-recdis-fields. Some inspection systems

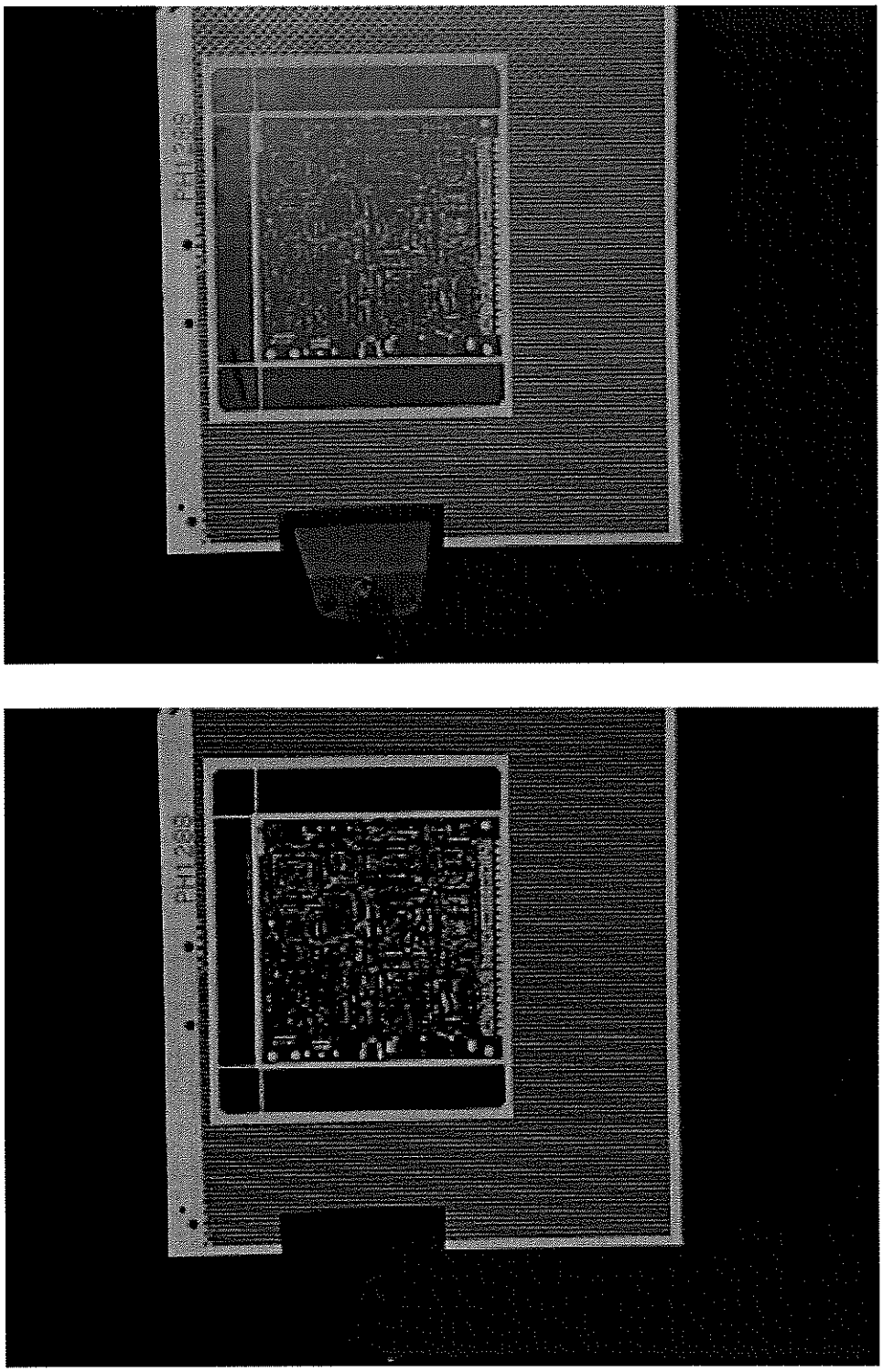


Figure 4.21: *Example of frontlighting (top) and backlighting (bottom).*

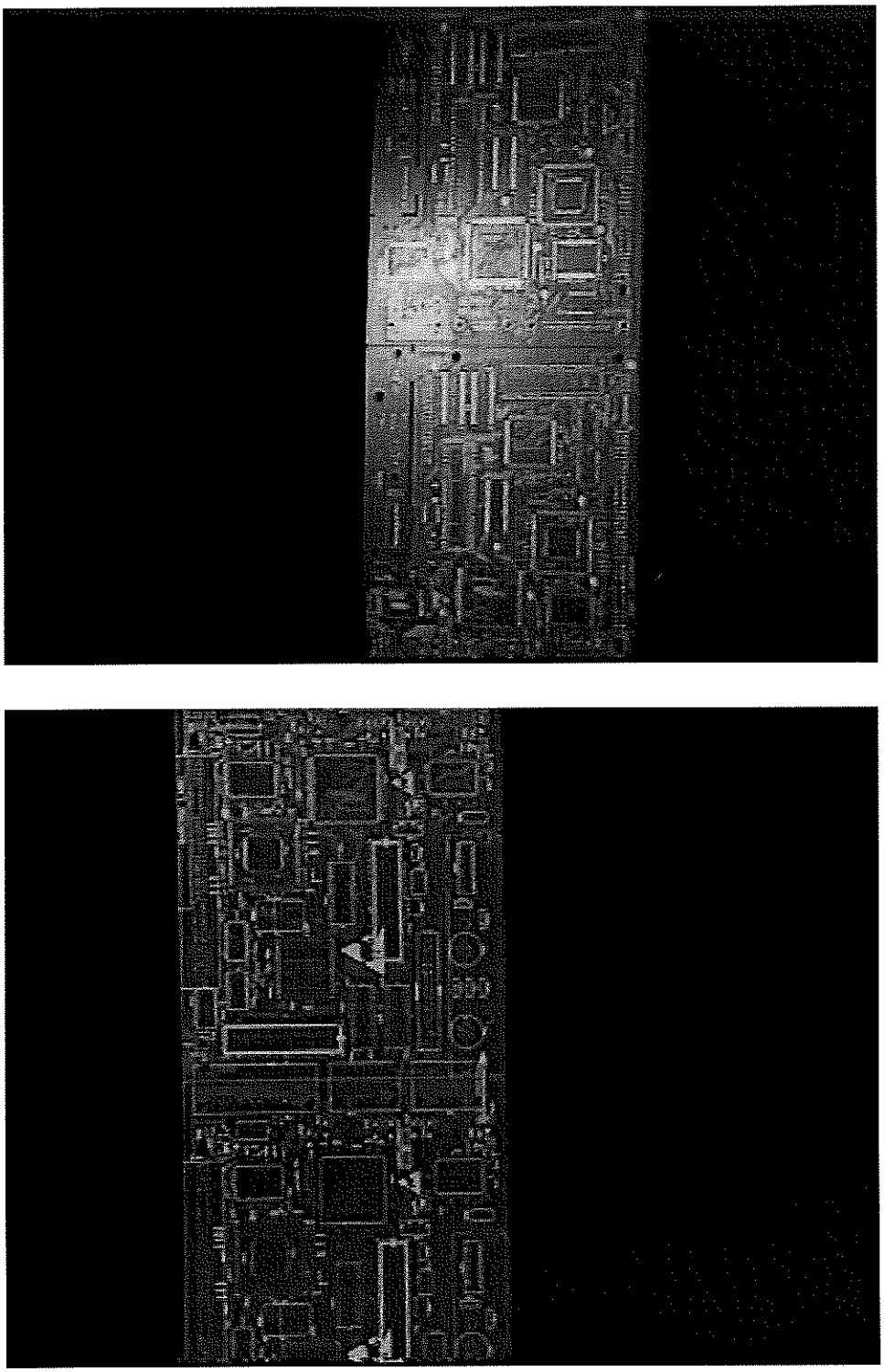


Figure 4.22: *Directional lighting (top) yields large changes in brightness due to specular reflections. These differences can be reduced through diffuse (omni-directional) lighting (bottom).*

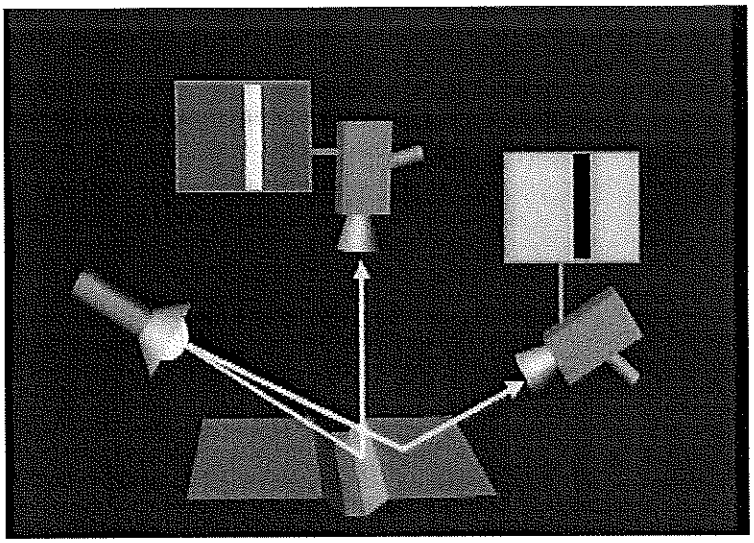


Figure 4.23: *Looking at a surface under bright or dark field conditions may facilitate the detection of flaws. The groove represents a crack in the surface. Contrast between the crack and the rest of the surface can be maximised by choosing an appropriate viewing direction.*

will use both fields. They might use the bright field for surface imperfections like cracks and the dark field for the detection of defects in the surface plane such as stains.

Fig. 4.24 shows the effect of choosing the appropriate angle for visualising a scratch in a metal surface. The crack yields a bright streak when viewed far enough into the dark field (bottom image). It is nearly invisible when viewed close to the bright field (top).

Directional lighting can also be used to generate sharp shadows in order to find clear object silhouettes. The shadows casted on the background by the object outlines yield information about its shape (shape from shadow) as do the gray levels of the object surface (shape from shading) if the surface reflectance properties and the positions of the light sources are known. The latter is still a subject that is largely of academic interest only, however.

Polarized lighting

As already discussed in section 2.1.2, light can be represented by transverse waves. When the electric field vector remains withing a fixed plane, the light is said to be linearly polarized. Section 2.2.1 already explained how polarization is affected by reflection and this influence depends on the nature of the reflecting material. *Polarized lighting* uses polarized light to exploit this material dependence.

Before the actual illumination techniques are discussed, it is important to briefly discuss the effect that polarization filters have. A typical setup will use two filters that only transmit light that is polarized in one particular direction, oriented parallel to the so-called *transmission axis*. One of the filters is placed after the light source to linearly polarize the light, whereas the second is placed in front of the sensor. The relative orientations of

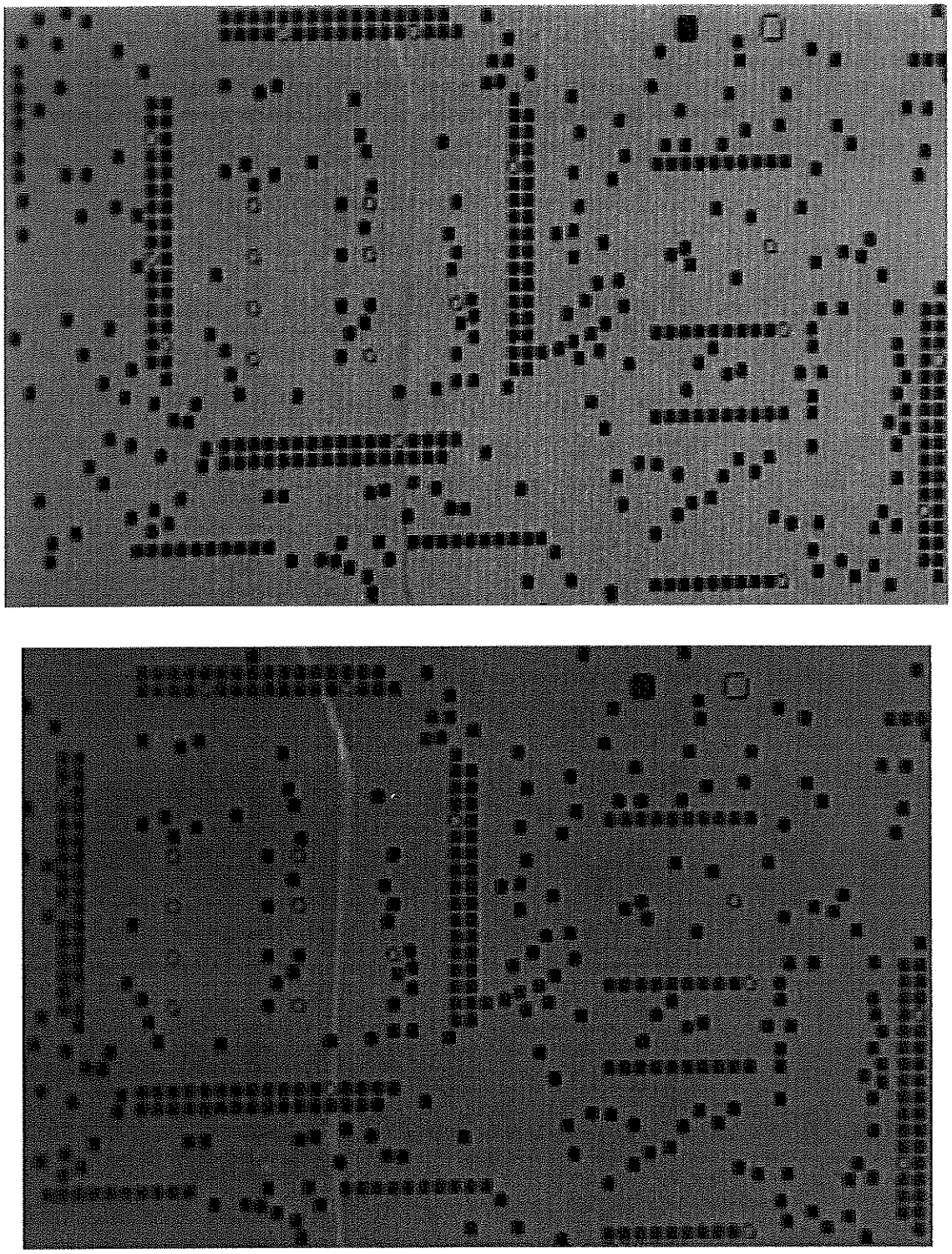


Figure 4.24: *If viewed under an inappropriate angle with respect to the angle of illuminiation the crack is hardly visible (top), whereas moving far enough into the dark field (bottom) makes the crack visible.*

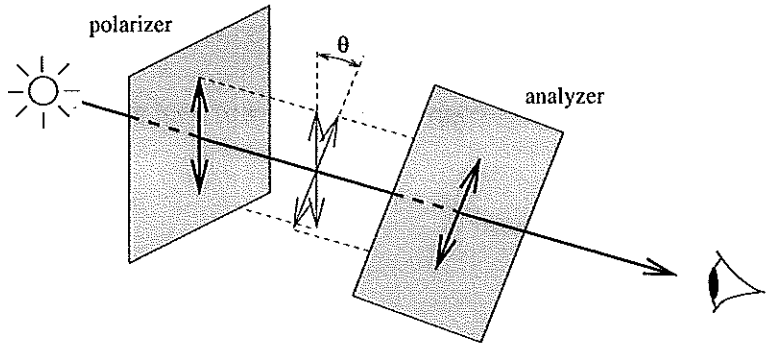


Figure 4.25: *Two polarization filters in cascade, with their relative transmission axes rotated over an angle θ.*

the transmission axes with respect to each other and the plane of incidence on the surface under scrutiny determine the irradiance at the position of the camera. First, consider what happens when two subsequent polarization filters are not aligned, as depicted in fig. 4.25. The first filter that polarizes the light is usually referred at as the *polarizer*, the filter before the sensor is called the *analyzer*. If the polarizer transmits a wave with amplitude $\|\mathbf{E}_0\|$, then only its component parallel to the transmission axis of the analyzer will pass that second filter. This component is given by $\|\mathbf{E}_0\| \cos\theta$ and using eq. 2.1, it follows that the irradiance at the sensor is

$$I(\theta) = \frac{1}{2}\sqrt{\epsilon\mu}\|\mathbf{E}_0\|^2 \cos^2\theta \ .$$

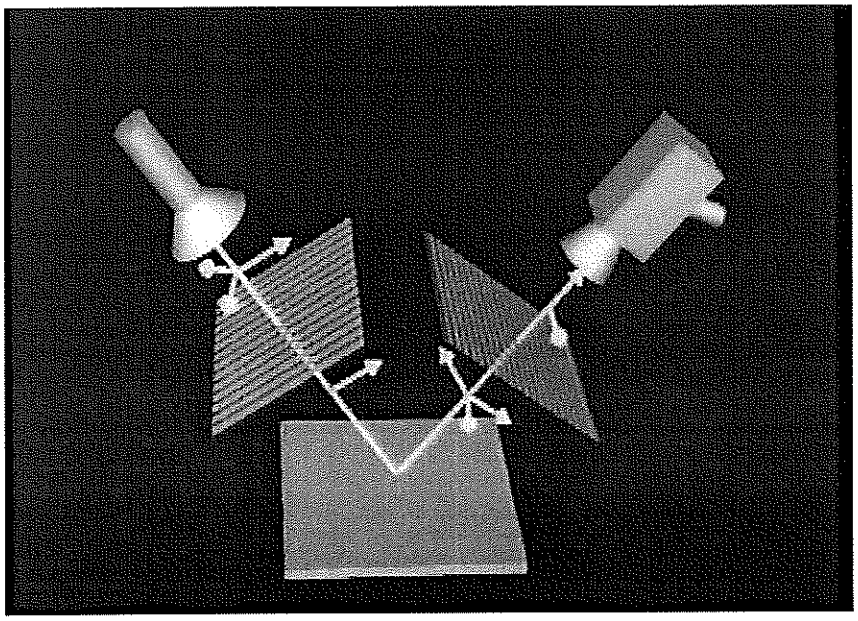
In fact, this is an oversimplification. Even for $\theta = 0$ polarization filters will reduce the transmitted irradiance. Therefore, a more appropriate formulation is

$$I(\theta) = I(0)\cos^2\theta \ ,$$

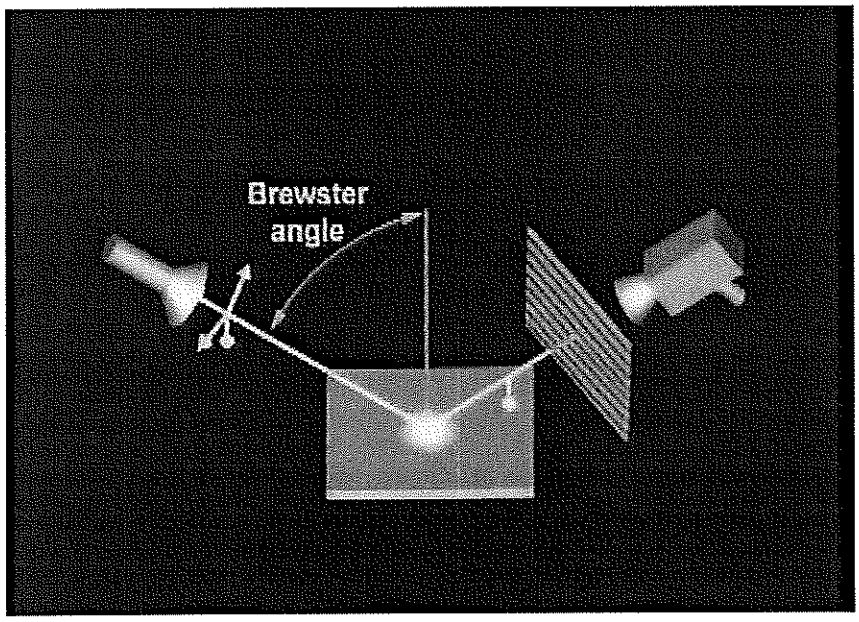
where $I(0)$ is the irradiance at the sensor for $\theta = 0$. This is *Malus's law*. Notice that if the filters are *crossed*, i.e. $\theta = 90^o$, then all light is blocked.

Using polarized light then follows one of two strategies, depending on the task. The first approach helps to remove specularities or to distinguish diffusely reflecting parts from specularly reflecting parts. The second approach helps to distinguish dielectrics from metals. The setups corresponding to these two cases are shown in fig. 4.26. Both configurations are now discussed and illustrated.

With the polarizer and analyzer crossed as in case a of the figure, specular reflection can be removed. For the shown setup, specular reflection will only consist of the parallel component, as indicated in fig. 2.8. This is eliminated by the analyzer. Hence, the sensor will mostly capture diffusely reflected light. Using such a setup, the specularly reflecting metal strips on a PCB board, which in a normal image without polarization filters will show high intensity specularities at places, can be made to look homogeneously darker than the diffusely reflecting board. This way, problems with glare from metal surfaces can be alleviated. Fig. 4.27 illustrates the principle. In the top image, taken without polarisation filters, the diffusely reflecting dielectric substrate of the plate and its specularly reflecting



a: diffuse / specular



b: dielectric / metal

Figure 4.26: *Two basic configurations using polarisation, one for removing specularities or distinguishing specularly reflecting from diffusely reflecting parts, the other for distinguishing metal parts from dielectric parts.*

metal covering change relative contrast at different places. Note for instance that the holes in the metal are darker, whereas the metal is bordered by a brighter portion of the dielectric substrate. The bottom image shows the use of the crossed polarizer/analyzer setup, yielding darker metal everywhere.

In the second case, the reflecting surface is made to act as a polariser in the case of dielectric surfaces (case **b** in fig. 4.26). One makes sure that the incident angle θ_i and the angle of viewing θ_r are the *Brewster angle*, i.e. the angle where the reflectance for the parallel component vanishes for dielectrics (see fig. 2.8). This component is eliminated by the analyser, which has to be given the orientation of 4.27. Whereas light reflected by dielectric parts of the surface is blocked, that coming from a metal part reaches the sensor, as for metals both the parallel and the orthogonal components are reflected (again, see fig. 2.8). Fig. 4.28 illustrates a result obtained with this approach. Both the PCB dielectric and the metal contacts and paths reflect specularly in this case (the dielectric is covered by a shiny translucent plastic layer). The top image is taken without a polarisation filter, the bottom image with the setup as described. The metal parts stand out as bright patches.

It should be kept in mind when using polarization filters, that a serious reduction in light power results. For the kind of setups in this section less than 20% of the light power initially captured by the polarizer may actually be available at the sensor.

Coloured lighting

Coloured-lighting can be useful to highlight regions of a similar colour. In combination with an appropriate band-pass filter all light but that of a monochromatic light pattern can be attenuated.

Fig. 4.29 gives an example. To find the (red) label, coloured (red) light was used (bottom image in fig. 4.29). As a result, the green PCB substrate material is darker than when using white light (top image in fig. 4.29).

Diffuse and specular reflection will often behave differently with respect to colour as well. Dielectric materials for instance, have specular reflections with the same colour of the light source, and diffuse reflection with some wavelengths eliminated.

When colours should be compared as with inspection tasks, the light sources must have a carefully calibrated spectral composition. The colour of two surfaces may look the same under one type of illumination and quite different with another. It e.g. happens that colours of cloths seem to match quite well while standing in a shop, but no longer once outside. The change from the artificial lighting to daylight brings different wavelengths to the fore. This is an illustration of *metamerism*. This term is given two related, but different meanings in the literature. First, two surfaces are said to be metameric if their colours match, despite differences in their spectral reflectance. As a second meaning, two surfaces are said to be the more metameric, the more their apparent colours change with illumination.

It is also important that the spectral sensitivity function of the sensors be known. A scene may look quite different to a camera than to the human observer.

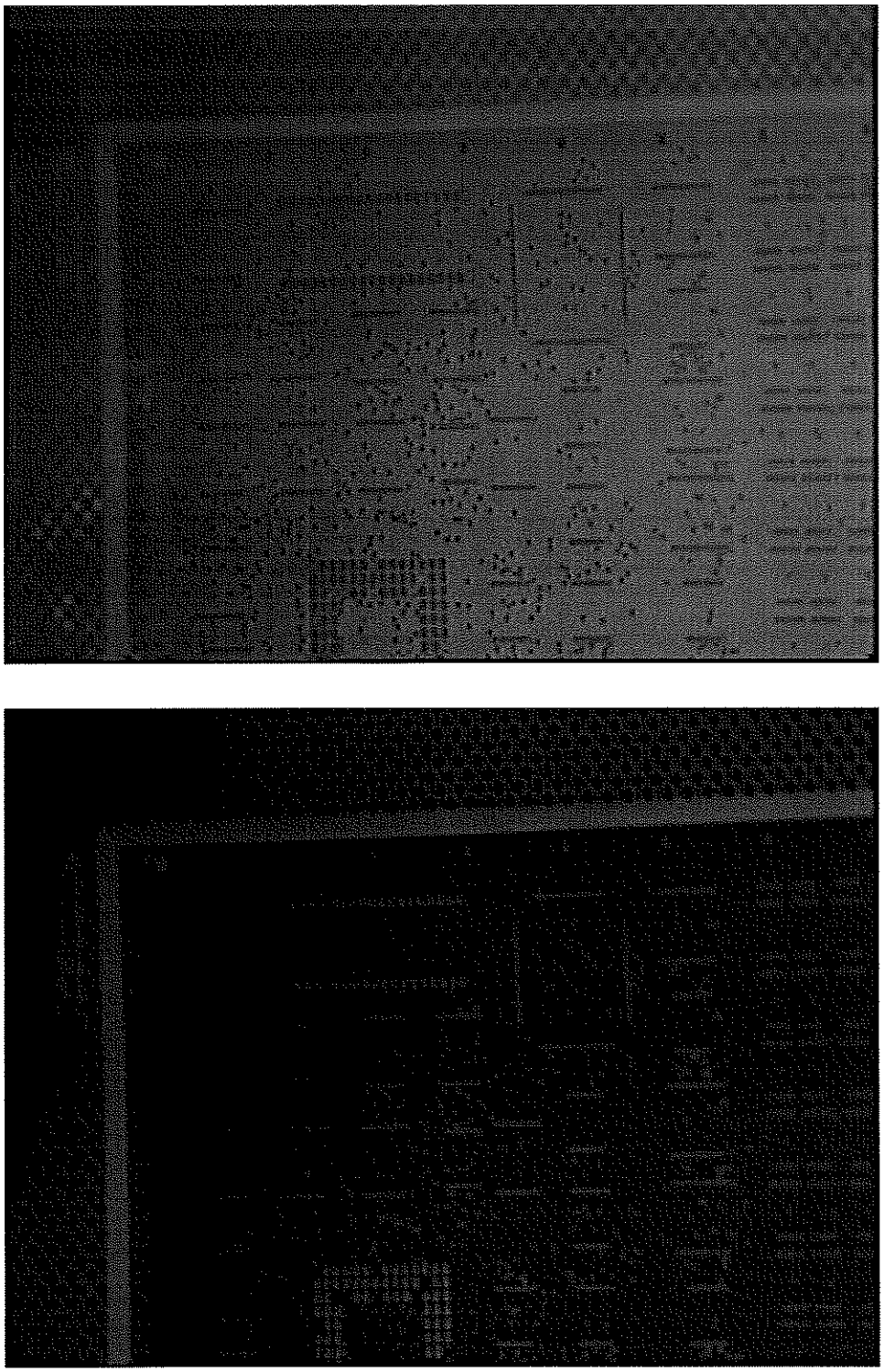


Figure 4.27: *The top image shows acquisition without the use of polarisation, whereas the bottom image shows the result of a crossed polarizer/analyzer setup. Note how the specularly reflecting metal is now darker than the diffusely reflecting dielectric substrate everywhere.*

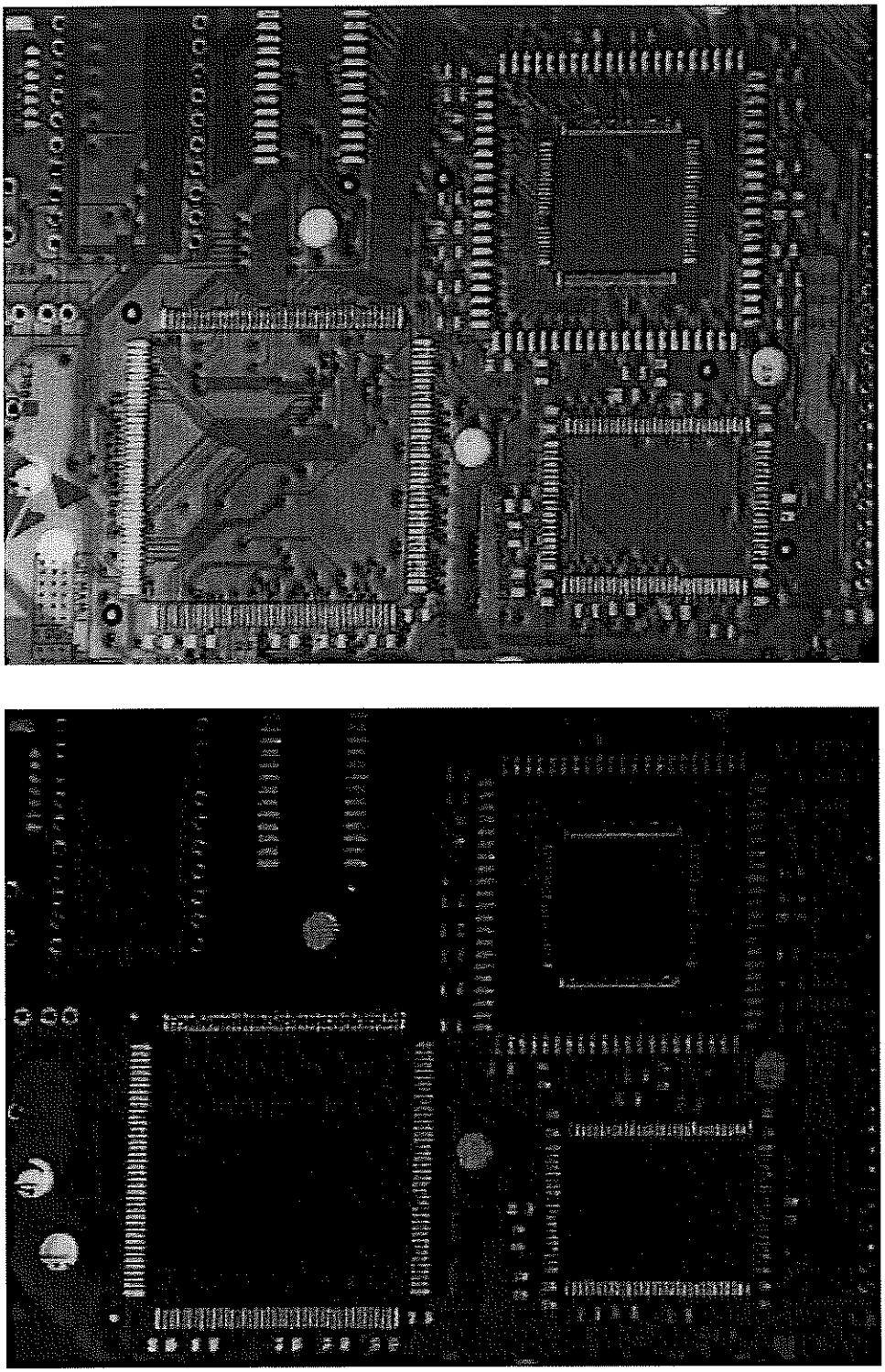


Figure 4.28: *Both the dielectric plate and the metal conductors reflect specularly in this example. Without the use of the polarisation filter it would be hard to distinguish the metal paths from the background. The use of the setup with polarisation filter (bottom image) facilitates making this distinction.*

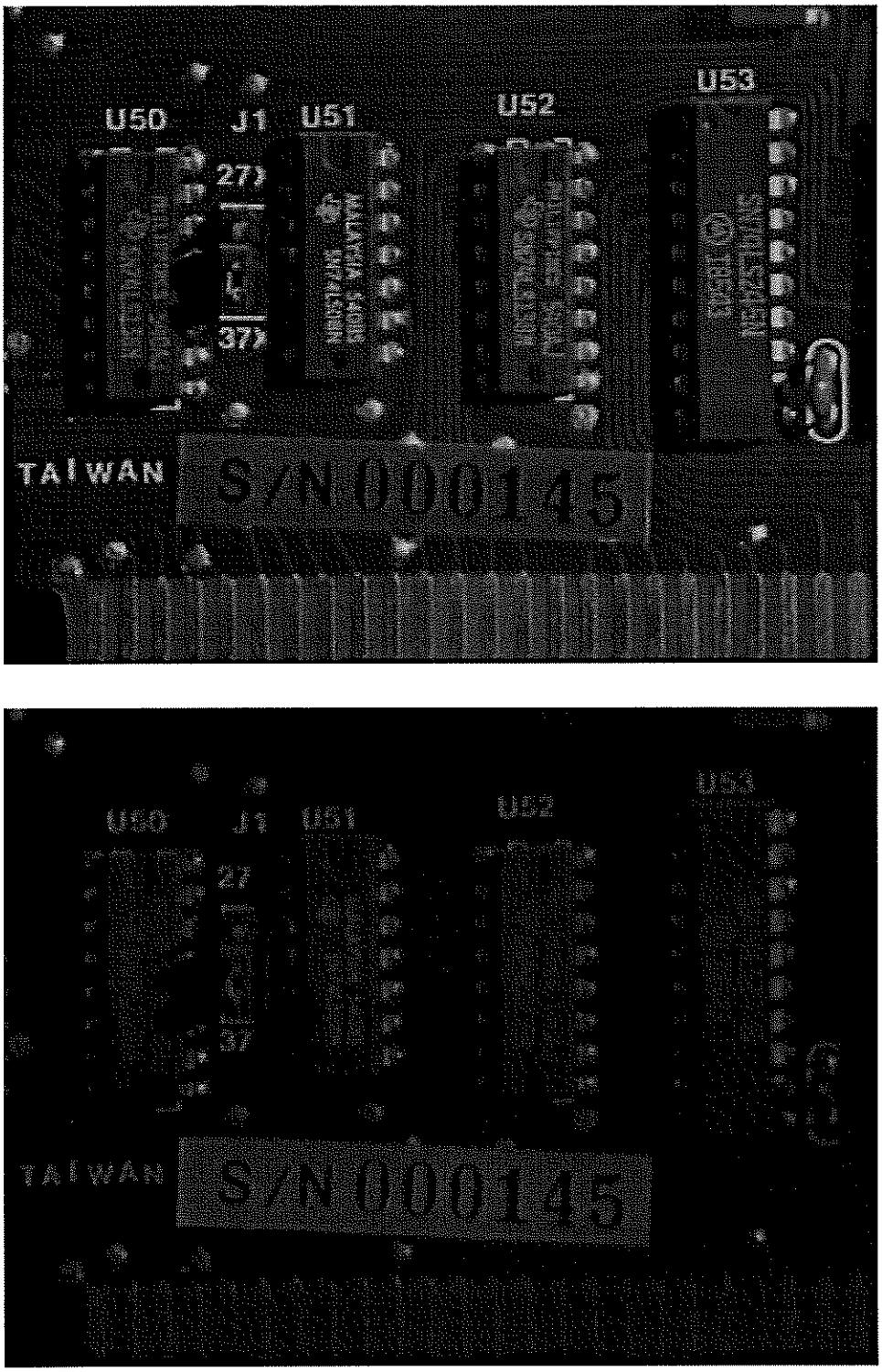


Figure 4.29: *The contrast of the label at the bottom of the PCB is increased upon replacing white light (top) by light of the label's colour (bottom).*

Structured lighting

Structured-lighting modulates the light pattern in space and/or time. The corresponding light patterns are often produced by lasers, which produce a very narrow beam of high intensity. This high intensity facilitates the detection of the pattern in the presence of ambient light, and so does the monochromatic spectrum of laser light. Objects will distort the projected pattern when viewed through the sensor. These distortions provide range or slant and tilt information. More about this is to follow (see chapter 11), but fig. 4.30 already illustrates the illumination aspect of the matter. The top image shows a PCB of which one might e.g. want to check whether chips have really been placed at all the position where they are supposed to be. One could check this by projecting lines. Depending on where they project in the image, the chips are present or missing. The technique is often combined with a bandpass filter that only lets the wavelength of the pattern light pass. This leads to a better visibility of the pattern (bottom image).

Stroboscopic lighting

Stroboscopic lighting makes use of a high intensity, short duration lamp flash. It is important when object motion is fast compared to the image formation (integration or shutter opening) time, which typically is 20 to 30 ms. In that case image blurring might result. For CCD cameras with adaptable integration time, increasing light intensity and decreasing integration time can be sufficient. But vidicons and non-adaptable CCD's need stroboscopic lighting as an artificial regulation of frame acquisition time. Good synchronization of object presentation, strobe firing, and image formation must be provided. Some cameras have special arrangements for this purpose. Fig. 4.31 gives an example of the effect of stroboscopic lighting.

4.5.3 An example

In order to fully convince the reader of the simplification in processing that a carefully studied illumination technique can bring, a practical example is discussed. The task is to inspect solder joints on PCBs (Printed Circuit Boards). In particular, joints with insufficient solder have to be identified. Examples for a good and a bad joint are schematically drawn in fig. 4.32. Using normal lighting would result in an image like that of fig. 4.33. This image contains good and bad joints. As the figure shows, these are extremely difficult to tell apart. Changing the illumination setup appropriately can simplify the problem drastically. Fig. 4.34 shows a modified camera-light configuration that prevents obliquely reflected light from entering the camera lens. By blocking this light, only light reflected by ill-shaped joints can reach the camera, as illustrated in the figure. Notice that the setup is shielded from ambient light. The shield was black, minimizing repeated reflections on the inside. A diffusor equally spread the light coming from a circular fluorescent tube placed around the camera lens. As a result, bad joints show up as bright spots. Image processing is reduced to the detection of such spots. The image of the same scene but with this

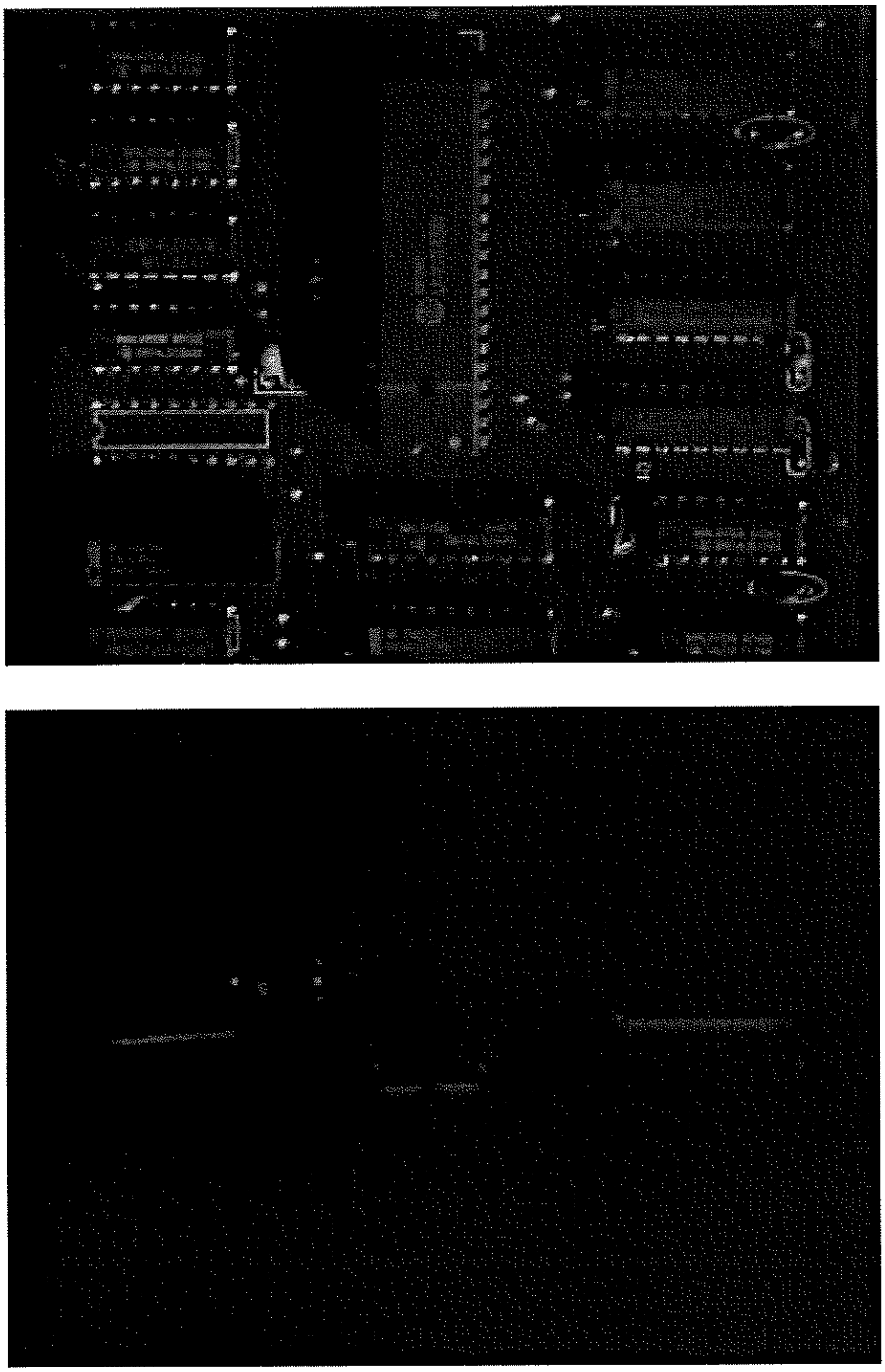


Figure 4.30: *As explained in chapter 11, projecting patterns on scenes can simplify the extraction of three-dimensional information. In combination with a bandpass filter for the colour of the pattern, it clearly stands out.*

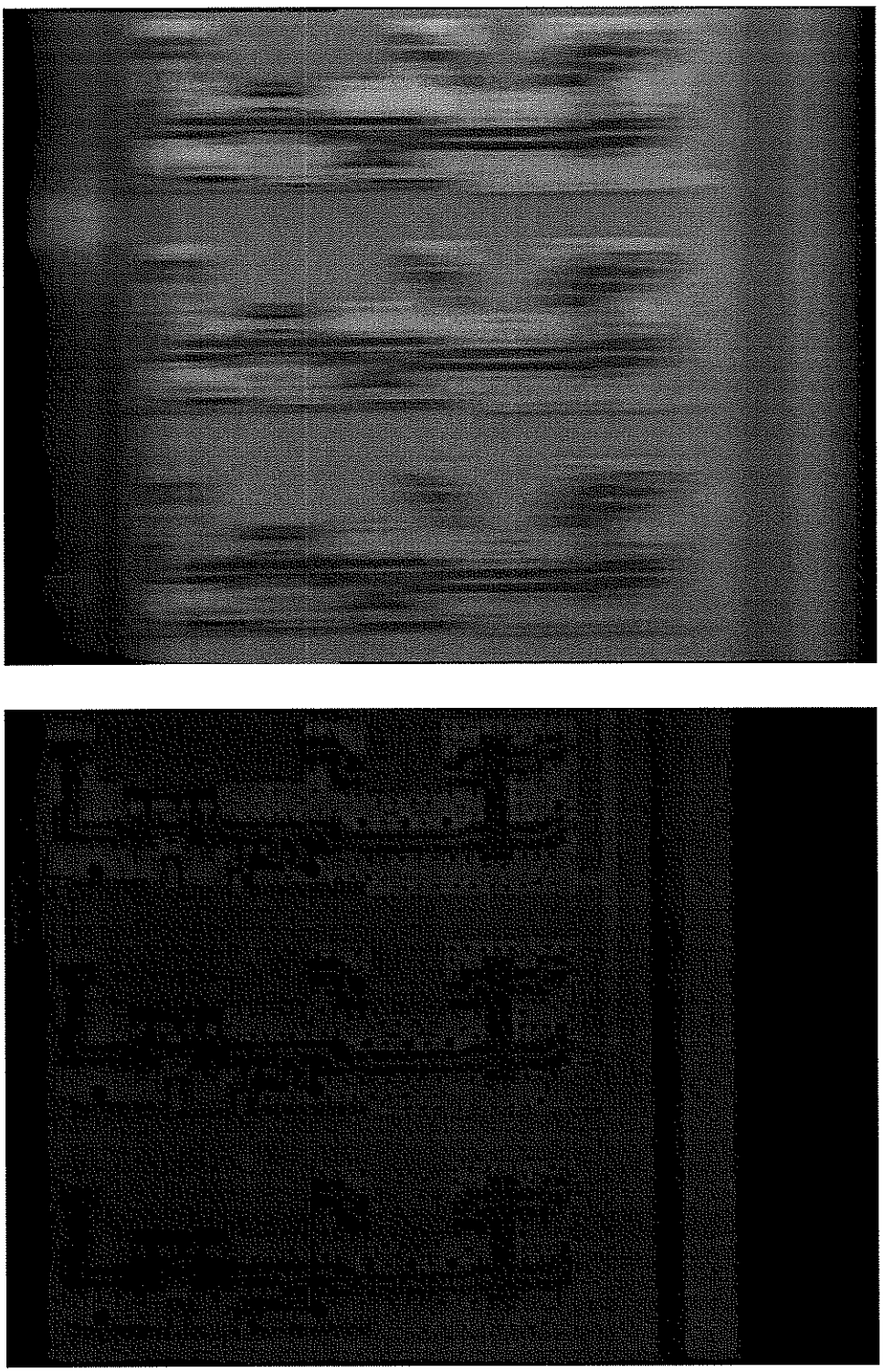


Figure 4.31: *Stroboscopic light can be used to eliminate motion blur (top image) by providing for an artificially shortened sensor integration time.*

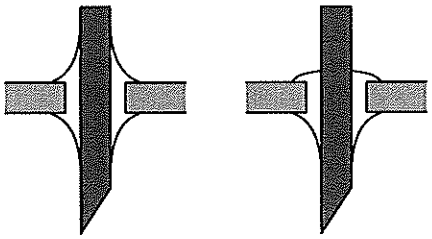


Figure 4.32: *Left: good joint, right: bad joint.*

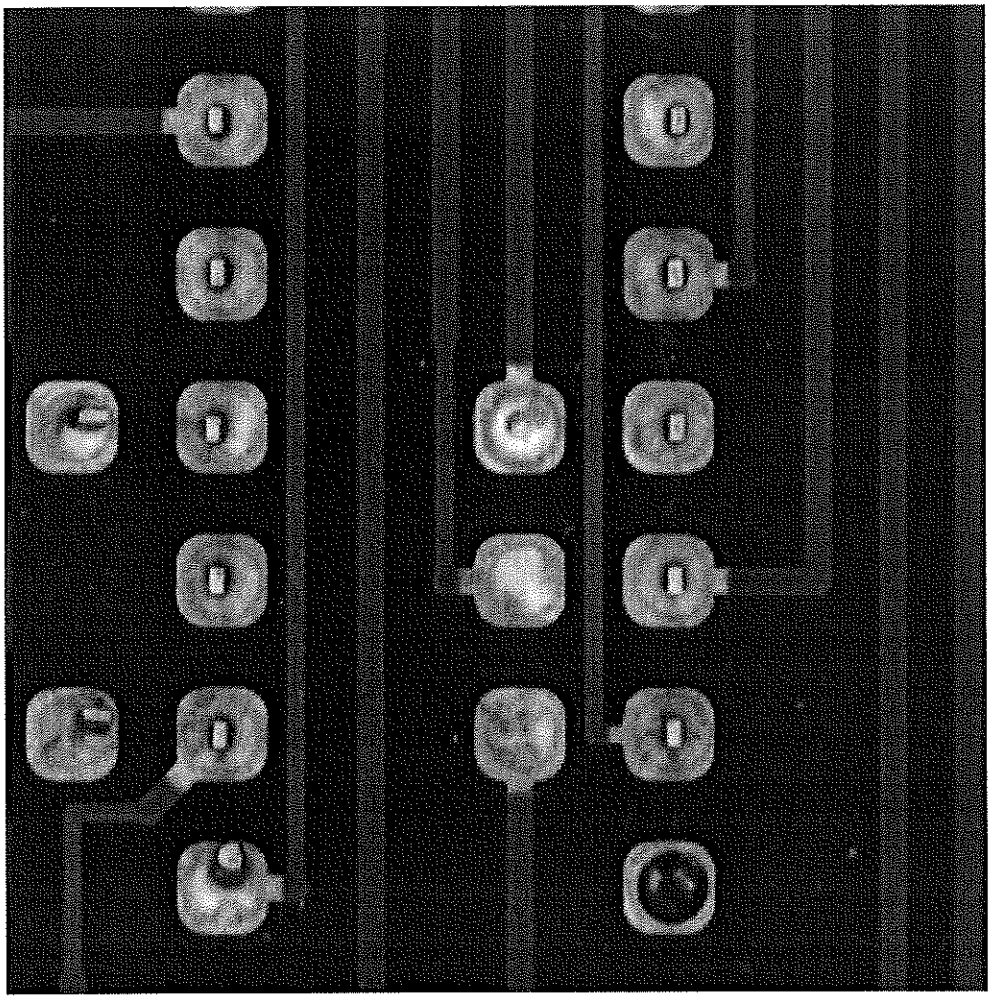


Figure 4.33: *With normal indoor illumination bad joints can hardly be distinguished from good ones.*

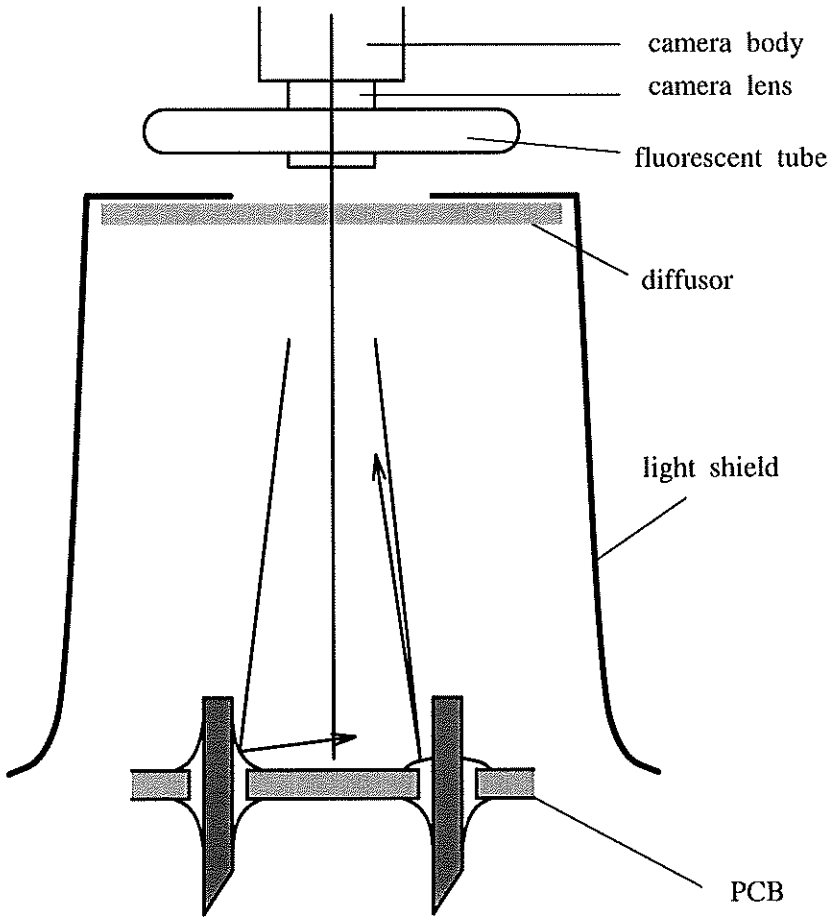


Figure 4.34: *The controlled lighting setup prevents ambient light from influencing the scene. A circular fluorescent lamp is the only light source. With a good joint, all light will be reflected in directions blocked by the top aperture. Bad joints, however, will reflect part of the light towards the camera aperture.*

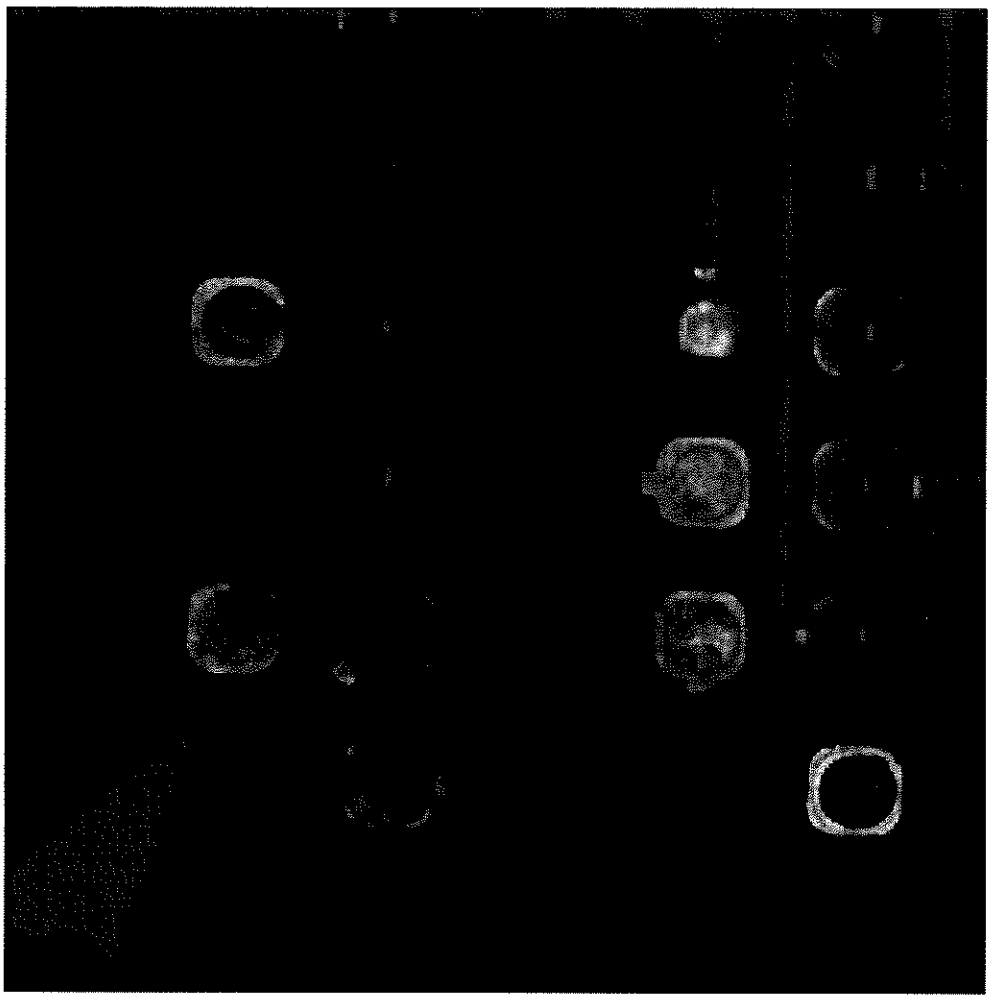


Figure 4.35: *Using the lighting setup of fig. 4.34 the bad joints are highlighted as bright spots.*

controlled illumination is shown in fig. 4.35.

4.6 II - Alternative types of image formation

4.6.1 Log-polar images

4.6.2 Panoramic vision

4.6.3 Medical scanners

4.7 Further reading

not available yet