Real Cameras and their Calibration

Carlo Tomasi

March 22, 2021

Real cameras behave somewhat differently from pinhole cameras. Since 3D reconstruction is an ill-posed computation[1], these differences need to be accounted for when working with real cameras. The next Section describes the key discrepancies, and Section 2 then shows how they can be addressed through a process called *camera calibration*.

1 Real Cameras

The pinhole camera is a useful and simple reference system for talking about the geometry of image formation. However, this device has a fundamental problem: If the pinhole is large, the image is blurred, and if it is small, the image is dim. When the diameter of the pinhole tends to zero, the image vanishes.[2] For this reason, lenses are used instead. Ideally, a lens gathers a whole cone of light from every point of a visible surface, and refocuses this cone onto a single point on the sensor plane. Unfortunately, lenses only approximate the geometry of a pinhole camera. The most obvious discrepancies relate to focusing and distortion, and are described in the next two Sections.

An array of photo-sensitive devices called *pixels* then measures the light on the sensor plane. What happens in the pixels and after them is called *sensing*, and is a matter of electronics. Key aspects of sensing, including how cameras handle color, are outlined in Appendix A.

1.1 Focusing

Figure 1 (a) illustrates the geometry of image focus. In front of the camera lens[3] there is a circular diaphragm of adjustable diameter called the *aperture*. This aperture determines the width of the cone of rays that hits the lens from any given point in the world.

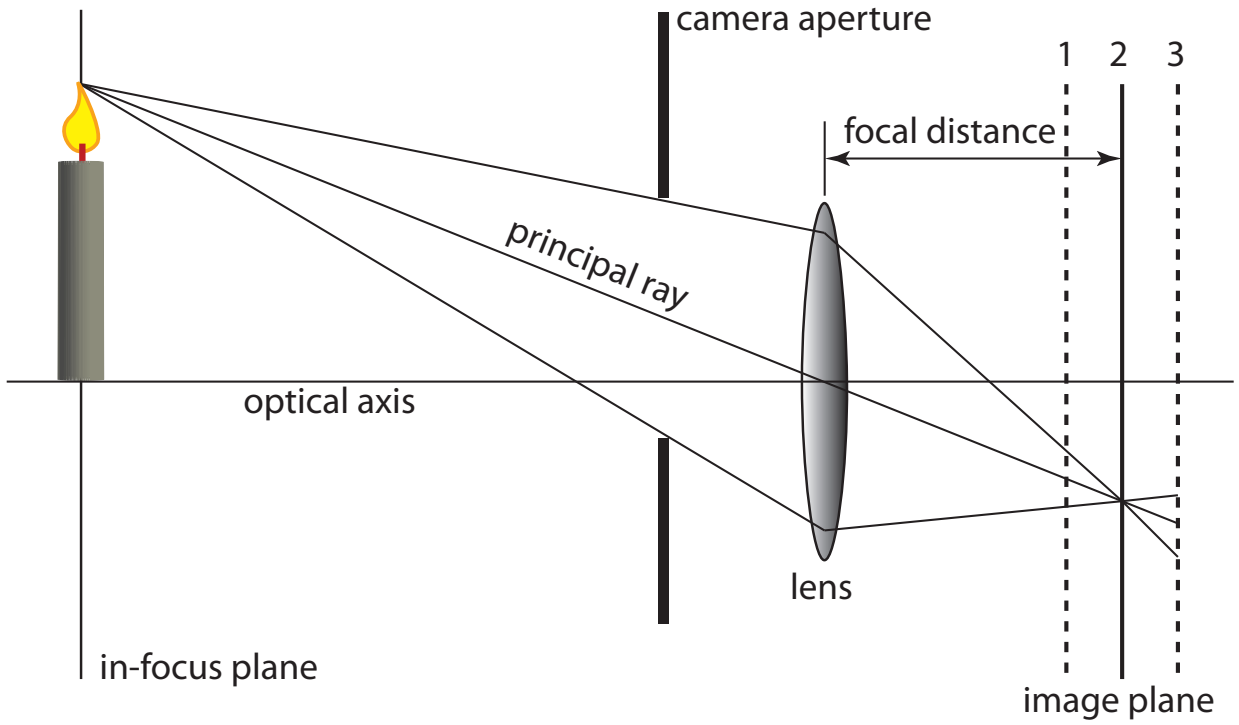
Consider for instance the tip of the candle flame in the Figure. For the image of the candle tip to be sharply focused, the lens must funnel all of the rays from the candle tip that make it through the aperture onto a single point on the image plane.

This condition is achieved by changing the focal distance, that is, the distance between the lens and the image plane. By studying the optics of light diffraction through the lens, it can be shown that the further the point in the world, the shorter the focal distance must be for sharp focusing: An image plane in position 1 in the Figure would focus points that are farther away than the candle, and an image plane in position 3 would focus points that are closer by.

[1] A computation is *ill-posed* when small perturbations of the input cause large changes in the output.

[2] In reality, blurring cannot be reduced at will, because of diffraction limits.

[3] Or inside the block of lenses, depending on various factors.



(a)

(b)

(c)

Figure 1: (a) If the image plane is at the correct focal distance (2), the lens focuses the entire cone of rays that the aperture allows through the lens onto a single point on the image plane. If the image plane is either too close (1) or too far (3) from the lens, the cone of rays from the candle tip intersects the image in a small ellipse (approximately a circle), producing a blurred image of the candle tip. (b) Image taken with a large aperture. Only a shallow range of depths is in focus. (c) Image taken with a small aperture. Everything is in focus.

Since the correct focal distance depends on the distance of the world point from the lens, for any fixed focal distance, only the points on a single plane in the world are in focus. If the image plane is at the wrong distance (cases 1 and 3 in the Figure), the cone of rays from the candle tip intersects the image plane in an ellipse, which for usual imaging geometries is very close to a circle. This is called the *circle of confusion* for that point. When every point in the world projects onto a circle of confusion, the image appears to be blurred.[4]

The dependence of focus on distance is visible in Figure 1(b): the lens was focused on the vertical, black and white stripe visible in the image, and the books that are closer are out of focus. The books that are farther away are out of focus as well, but by a lesser amount, since the effect of depth is not symmetric around the optimal focusing distance. Photographers say that the lens with the settings in Figure 1(b) has a *shallow depth of field*.

The depth of field can be increased, that is, the effects of poor focusing can be reduced, by making the lens aperture smaller. As a result, the cone of rays that hit the lens from any given point in the world becomes narrower, the circle of confusion becomes smaller, and the image becomes more sharply focused everywhere. This can be seen by comparing Figures 1 (b) and (c). Image (b) was taken with the lens aperture opened at its greatest diameter, resulting in a shallow depth of field. Image (c), on the other hand, was taken with the aperture closed down as much as possible for the given lens, resulting in a much greater depth of field: all books are in focus to the human eye. The price paid for a sharper image was exposure time: a small aperture lets little light through, so the imaging sensor had to be exposed longer to the incoming light: 1/8 of a second for image (b) and 5 seconds, forty times as long, for image (c).

The focal distance at which a given lens focuses objects at infinite distance from the camera is called the *rear focal length* of the lens, or *focal length* for short.[5] All distances are measured from the center of the lens and along the optical axis. Note that the focal length is a lens property, which is usually printed on the barrel of the lens. In contrast, the focal distance is the distance between lens and image plane that a photographer selects to place a certain plane of the world in focus. So the focal distance varies even for the same lens.[6]

In photography, the aperture is usually measured in *stops*, or f*-numbers*. For a focal length f, an aperture of diameter a is said to have an f-number

$$n = \frac{f}{a} ,$$

so a large aperture has a small f-number.[7] To remind one of this fact, apertures are often denoted with the notation f/n. For instance, the shallow depth of view image in Figure 1 (b) was obtained with a relatively wide aperture $a = f/4.2$ (a small f-number), while the greater depth of field of the image in Figure 1 (c) was achieved with a much narrower aperture $a = f/29$ (a large f-number).

Why use a wide aperture at all, if images can be made sharp with a small aperture? As was mentioned earlier, sharper images either are darker, or require longer exposure times. In the example above, the ratio between the *areas* of the apertures is $(29/4.2)^2 \approx 48$. This is more or less consistent with the fact that the sharper image required forty times the exposure of the blurrier one: 48 times the area means that the lens focuses 48 times as much light on any given small patch on the image, and the exposure time can be

[4] You should know enough about convolution by now to see that the blurred image is a convolution of the ideal, sharp image with a pillbox kernel. A more detailed treatment of lens optics, including diffraction effects, would show that to be an approximation.

[5] The *front focal length* is the converse: the distance to a world object that would be focused on an image plane at infinite distance from the lens.

[6] This has nothing to do with zooming. A zoom lens lets you change the focal length as well, that is, modify the optical properties of the lens.

[7] The terminology, while standard, is a bit confusing: f is the focal length, and the f-number is n.

Figure 2: A shallow depth of field draw attention to what is in focus, at the expense of what is not.

decreased accordingly by a factor of 48. So, wide apertures are required for subjects that move very fast (for instance, in sports photography). In these cases, long exposure times are not possible, as they would lead to *motion blur*, a blur of a different origin (motion in the world) than poor focusing. Wide apertures are often aesthetically desirable also for static subjects, as they attract attention to what is in focus at the expense of what is not, because of the resulting shallow depth of field. This is illustrated in Figure 2, initially downloaded from
`http://www.hp.com/united-states/consumer/digital_photography/take_better_photos/tips/depth.html` and no longer available there.

1.2 Distortion

Even the high quality lens[8] used for the images in Figure 1 exhibits distortion. For instance, if you place a ruler along the vertical edge of the blue book on the far left of the Figure, you will notice that the edge is not straight. Some curvature is visible also in the top shelf. This is geometric *pincushion distortion*. This type of distortion, illustrated in Figure 3(b), moves every point in the image away from the principal point, by an amount that increases with the square of the distance of the point from the principal point. The reverse type of distortion is called *barrel distortion*, and draws image points closer to the principal point by an amount that increases with the square of their distance from it. Because they move image points towards or away from the principal point, both types of distortion are called *radial*. While non-radial distortion does occur, it is typically negligible in reasonably good lenses, and is henceforth ignored.

Distortion can be quite substantial, either by design (such as in non-perspective lenses like fisheye lenses) or to keep the lens inexpensive and with a wide field of view. Accounting for distortion is crucial in computer vision algorithms that use cameras as measuring devices, for instance, to reconstruct the three-dimensional shape of objects from two or more images of them.

[8] Nikkor AF-S 18-135 zoom lens, used for both images (b) and (c).

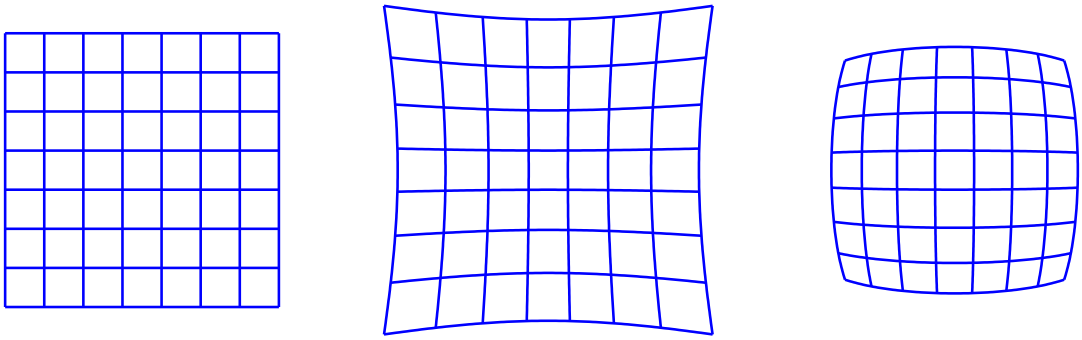


Figure 3: (a) An undistorted grid. (b) The grid in (a) with pincushion distortion. (c) The grid in (a) with barrel distortion.

Practical Aspects: Achieving Low Distortion. Low distortion can be obtained by mounting a lens designed for a large sensor onto a camera with a smaller sensor. The latter only sees the central portion of the field of view of the lens, where distortion is usually small.

For instance, lenses for the Nikon D200 camera used for Figure 1 are designed for a 23.6 by 15.8 millimeter sensor. Distortion is small but not negligible (see Figure 1 (c)) at the boundaries of the image when a sensor of this size is used. Distortion would be much smaller if the same lens were mounted onto a camera with what is called a "1/2 inch" sensor (which is really 6.4 by 4.8 millimeters in size), because the periphery of the lens would not be used. Lens manufacturers sell relatively inexpensive adaptors for this purpose. The real price paid for this reduction of distortion is a concomitant reduction of the camera's field of view (more on this in Appendix A).

2 Camera Calibration

Calibrating a camera means determining the parameters of the function that describes the mapping from the position of a point in the world to the position of the projection of that point on the image plane.[9] This mapping can be written as follows[10] for a pinhole camera with an image sensor (see Figure 4):

$$\begin{aligned} \mathbf{X} &= R(\mathbf{W} - \mathbf{t}) \\ \mathbf{x} &= p(\mathbf{X}) = \frac{1}{X_3} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \\ \boldsymbol{\xi} &= f\tilde{S}\mathbf{x} + \boldsymbol{\pi} \end{aligned} \tag{1}$$

where $\mathbf{W} = [W_1, W_2, W_3]^T$ is a world point in the world reference frame, $\mathbf{X} = [X_1, X_2, X_3]^T$ is $\mathbf{W}$ re-expressed in the camera's reference frame, $\mathbf{x} = [x_1, x_2]^T$ is the projection of $\mathbf{X}$ in the canonical image reference frame, and $\boldsymbol{\xi} = [\xi_1, \xi_2]^T$ is $\mathbf{x}$ re-expressed in the pixel image reference frame. The vector $\boldsymbol{\pi} = [\pi_1, \pi_2]^T$ is the principal point in image coordinates, f is the camera's focal distance, and the diagonal

[9] Assuming that this mapping exists implies that blurring caused by defocus is ignored. While a modest amount of blurring may cause uncertainty, the bias it produces is negligible.

[10] The notation used in these equations is somewhat simplified here. In particular, there is only one camera, so it would be unnecessarily cumbersome to use subscripts and superscripts for different reference frames.

elements of the matrix

$$\tilde{S} = \begin{bmatrix} \tilde{s}_1 & 0 \\ 0 & \tilde{s}_2 \end{bmatrix}$$

are the horizontal and vertical pixel scaling factors. The function p that relates world coordinates $\mathbf{X}$ to canonical image coordinates $\mathbf{x}$ is called the *canonical perspective projection* function, and refers to an ideal camera with unit focal distance. For mathematical convenience, coordinates are scaled by f in the transformation from $\mathbf{x}$ to $\boldsymbol{\xi}$, rather than during projection.

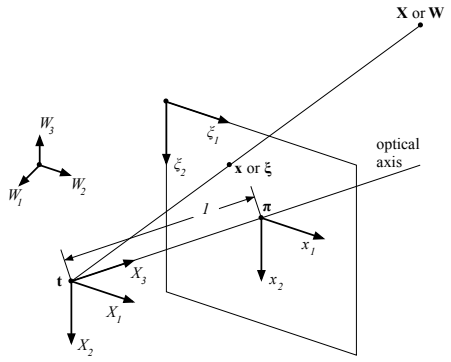


Figure 4: The main reference systems for a camera.

The two terms f and $\tilde{S}$ appear as a product in equation (1), and do not appear anywhere else. Because of this, they are defined only up to a scale factor: If f were multiplied by a nonzero constant and $\tilde{S}$ divided by the same constant, their product would not change. To account for this unavoidable ambiguity, equation (1) is rewritten as follows:

$$\boldsymbol{\xi} = S\mathbf{x} + \boldsymbol{\pi} \quad \text{where} \quad S = f\tilde{S} \tag{2}$$

and only S is computed during calibration. Whenever f or $\tilde{S}$ is known from camera specifications, the unknown quantity can be inferred from the known one and S.

In addition, most lenses also distort images, as we saw earlier, and Section 2.1 below introduces a simple mathematical model of lens distortion. The Section thereafter shows how to estimate the parameters of a camera model extended to include distortion.

2.1 Lens Distortion

High-quality lenses can be purchased for which distortion may be negligible. Some lenses, including high-quality ones, introduce distortion by design. These include fish-eye lenses, which are built to create wide panoramic images or special effects like in the image shown in Figure 5.

For these lenses, or for lenses that produce distorted images as a result of design compromises, one also needs to estimate the parameters of image distortion, which is an inherently nonlinear effect.

A camera model that includes distortion replaces equation (2) with the following two equations:

$$\begin{aligned} \mathbf{y} &= d(\mathbf{x}) \\ \boldsymbol{\xi} &= S\mathbf{y} + \boldsymbol{\pi} \end{aligned}$$

where d is the *distortion function*.

Figure 5: Paul Bourke took this image of downtown Perth, Australia, with a fisheye lens.
[Image from http://paulbourke.net/dome/cameras/]

Since lenses are radially symmetric around the optical axis, distortion is radially symmetric around the principal point of the image,[11] and this is why the distortion function d is most easily applied to the canonical coordinates $\mathbf{x}$, which are measured in a reference system whose origin is the principal point. For the same reason, d acts equally in all radial directions, and therefore takes the form of a scaling function

$$\mathbf{y} = d(\mathbf{x}) = \delta(r)\,\mathbf{x} \quad \text{where} \quad r = \|\mathbf{x}\| \ .$$

The function $\delta : \mathbb{R} \to \mathbb{R}$ is called the *radial distortion function* and depends only on the distance r of the undistorted point $\mathbf{x}$ from the principal point. Of course, the distortion function d is nonlinear, because of both the dependence of δ on the magnitude of $\mathbf{x}$ and the multiplication of $\delta(r)$ by $\mathbf{x}$.

The radial distortion function $\delta(r)$ is typically approximated by a low degree polynomial, whose coefficients are the model parameters to be estimated. It can be shown [1] that lens distortion must be an analytical function of $\mathbf{x}$, that is, it must be infinitely differentiable everywhere. This implies that when one sets, say, $\mathbf{x} = (x, 0)^T$, that is, when $\mathbf{x}$ is restricted to the x axis, the function

$$\delta(r(\mathbf{x})) = \delta(|x|)$$

must be infinitely differentiable everywhere. If δ is a polynomial, this implies that its odd coefficients must be zero, because the $2k+1$-st derivative of $|x|^{2k+1}$ is discontinuous at $x = 0$ for any nonnegative integer k, so the inclusion of odd powers would violate infinite differentiability at the origin. Thus, δ must have the following form:

$$\delta(r) = 1 + k_1 r^2 + k_2 r^4 + \dots \ .$$

Large powers r^{2k} are very flat close to the origin, so their effects on distortion are only noticeable close to the image boundaries. Because of this, few useful image measurements are typically available to constrain high-order coefficients, and it is common practice to only use terms up to r^4:

$$\delta(r) = 1 + k_1 r^2 + k_2 r^4 \ .$$

[11] Complex lenses, made of many different glass elements, can also exhibit tangential distortion, but this is typically negligible when compared to the radial component.

In summary, a camera model that includes distortion is as follows:

$$\begin{aligned}
\mathbf{X} &= R\,(\mathbf{W}-\mathbf{t}) \\
\mathbf{x} &= p(\mathbf{X}) \;=\; \frac{1}{X_3}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \\
\mathbf{y} &= d(\mathbf{x}) \;=\; \mathbf{x}\left(1+k_1\|\mathbf{x}\|^2+k_2\|\mathbf{x}\|^4\right) \\
\boldsymbol{\xi} &= S\mathbf{y}+\boldsymbol{\pi}\;.
\end{aligned}$$

This model can be viewed as a function

$$\boldsymbol{\xi} = \mathbf{c}(\mathbf{W};\mathbf{p}) \tag{3}$$

from $\mathbb{R}^3$ to $\mathbb{R}^2$ that depends on a set of parameters listed in the vector $\mathbf{p}$. The parameters $\boldsymbol{\pi}$, s_1, s_2, k_1, k_2 in the camera model are called *intrinsic parameters*, because they only depend on the camera internals, and not on the camera's position or orientation. Of course, if the lens is changed or zoomed in or out, or even focused to a different distance, these parameters change. So it is frequent in computer vision to use lenses whose settings can be mechanically locked.

The parameters in R and $\mathbf{t}$ are called the *extrinsic parameters*, as they depend on where the camera is in the world and on which way it is pointing, and these factors are external to the camera itself. The 9 entries of the rotation matrix R satisfy the six independent constraints implied by orthogonality:[12]

$$R^T R = I\;.$$

Because of these constraints, rotation can be represented succinctly by a vector $\mathbf{r}$ with three parameters. One way of doing so is shown in Appendix B, which also shows how to convert R to $\mathbf{r}$ and *vice versa*. Then, the vector $\mathbf{p}$ of parameters in equation (3) has twelve scalar entries:

$$\mathbf{p}^T = \left[\mathbf{r}^T, \mathbf{t}^T, \boldsymbol{\pi}^T, s_1, s_2, k_1, k_2\right]\;.$$

2.2 Calibration

Camera calibration typically proceeds as follows:

Setup: Construct a physical object and record the coordinates $\mathbf{W}_n$ of a set of N visible feature points on the object. These coordinates are measured in a system of reference attached to the object. It turns out that for calibration to yield a unique solution the points $\mathbf{W}_n$ cannot all be on the same plane.

Imaging: Take an image of this *calibration target* and record the image coordinates $\boldsymbol{\xi}_n$ of the feature points.

Optimization: Fit the parameter vector $\mathbf{p}$ to the measurements above by computing

$$\mathbf{p}^* = \arg\min_{\mathbf{p}} e(\mathbf{p}) \quad \text{where} \quad e(\mathbf{p}) = \frac{1}{N}\sum_{n=1}^{N} \|\mathbf{c}(\mathbf{W}_n;\mathbf{p}) - \boldsymbol{\xi}_n\|^2\;. \tag{4}$$

These steps are described next. The part on optimization loosely follows the lines of a popular article on the topic [3].

[12] The matrices on the two sides of this equation are symmetric, so there are only six independent scalar equations rather than 9.

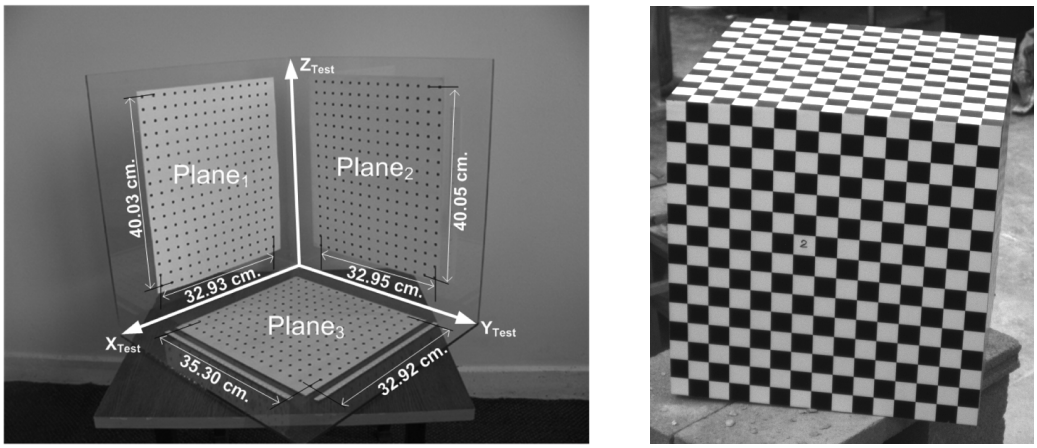


Figure 6: Two calibration targets.
[Left image from http://www.mdpi.com/1424-8220/9/6/4572/htm. Right image: Duke Computer Vision Lab.]

Setup Figure 6 shows two calibration targets. Good targets are made of aluminum or acrylic plastic for structural rigidity and accurate machining. The features are printed with a high-quality printer[13] onto a sheet of paper or plastic that is glued to the target. Alternatively, the features are silk-screened directly onto the target for even greater positional accuracy.

The features on the target on the left in Figure 6 are small, dark circles, and the features are the centers of the circles. The target on the right uses a checkerboard, and the features are the corners of the cells. The origin of the world reference system is a designated corner of the target, and its axes are along the edges of the target. The pitch and position of either grid is known, and therefore one can calculate the precise coordinates $\mathbf{W}_n$ of each feature (circle center or cell corner) in a predefined order and store these coordinates in a file.

Imaging Image coordinates can be measured by hand by looking at the image through an image browser, clicking on the features, and recording row (ξ_2) and column (ξ_1) coordinates.

Circles as features are not an ideal choice, because a circle projects to an ellipse in the image, and the center of the circle does not in general project to the center of the ellipse. This discrepancy is illustrated in Figure 7. The cell corners of a checkerboard pattern, on the other hand, are well defined.

Automatic methods for finding features have been developed as well. With a checkerboard pattern, the software finds the lines that delimit the rows and columns of the pattern, and determines cell corner coordinates as intersections between lines. If there is image distortion, a more accurate method is to fit quadratic curves to the boundaries between rows and columns, because these boundaries may be slightly curved.

A numerical optimization procedure may be used to find the line or curve parameters that maximize the integral of image gradient magnitude along the line or curve. This procedure is often initialized manually, by placing the line endpoints in approximately their correct positions with a cursor in an image browser. To save time, automatic initialization methods may be warranted if several cameras need to be calibrated, or

[13] Horizontal dimensions of points printed with a laser printer are highly accurate, because they depend on a precisely positioned laser beam. Vertical dimensions, on the other hand, depend on the accuracy of the rotation speed of the printer's drum. This accuracy is high only in good printers.

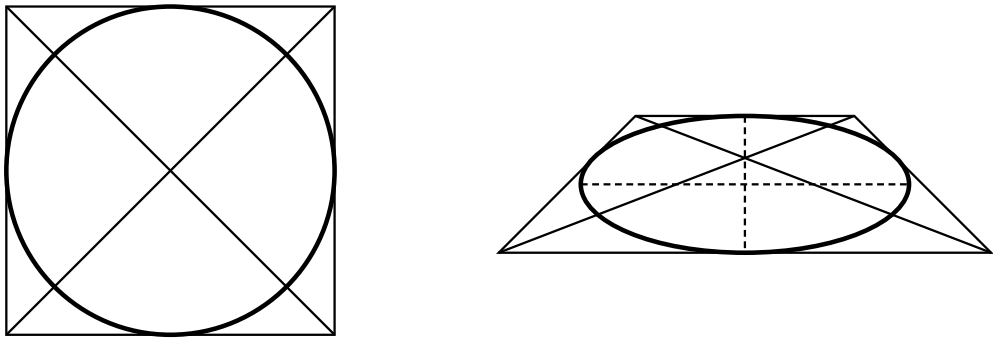


Figure 7: *Left:* The center of the circle is the intersection of the two solid diagonal lines. *Right:* The projection of the center of the circle is still the intersection of the two solid diagonal lines, because perspective projection preserves co-linearity. However, the circle projects onto an ellipse, and the center of the ellipse (intersection of the two dashed lines) is generally not where the two solid diagonals intersect.

if the same camera needs to be repeatedly re-calibrated when its position, orientation, or lens parameters change over time.

In any case, the image feature coordinates $\boldsymbol{\xi}_n$ are stored in a file in the same order in which the feature world coordinates $\mathbf{W}_n$ were stored. Care is needed to ensure that the correspondence between points in these two files is correct.

Optimization The error function $e(\mathbf{p})$ defined in equation (4) is generally non-convex. To use a local optimization method, it is therefore necessary to initialize the search for a minimum with a vector $\mathbf{p}_0$ of parameters that are close to the correct solution, lest a spurious local minimum is found.

The vector $\mathbf{p}_0$ is typically computed by solving an approximate version of problem (4) in which the distortion coefficients k_1 and k_2 are clamped to zero. In this way, the only remaining nonlinearity in the camera model is the projection function $p(\mathbf{X})$. As shown in Appendix C, a solution $\mathbf{p}_0$ to the simplified problem can be found by solving a homogeneous system of linear equations obtained by algebraic manipulation of the camera model. This method minimizes the algebraic Least-Squares error for the resulting linear system rather than the error function $e(\mathbf{p})$, so $\mathbf{p}_0$ is an approximation for two distinct reasons: Distortion is ignored, and an error function different from $e(\mathbf{p})$ is minimized. In the absence of distortion and image measurement errors, $\mathbf{p}_0$ and $\mathbf{p}^*$ would be the same. In reality, they are not, and a standard local minimization method is then used on the general camera model (with k_1 and k_2 now free to vary) to find $\mathbf{p}^*$ starting from $\mathbf{p}_0$.

2.3 Canonical Image Coordinates

After calibration, it is often useful in various applications[14] to convert image pixel coordinates $\boldsymbol{\xi}$ to canonical image coordinates $\mathbf{x}$. To this end, one needs to invert the distortion function $d(\mathbf{x})$. This can be accomplished by solving the following polynomial equation in $\mathbf{x}$ using a numerical root-finding algorithm such as Brent's method [2]:

$$d(\mathbf{x}) = \mathbf{y} \quad \text{where} \quad \mathbf{y} = S^{-1}(\boldsymbol{\xi} - \boldsymbol{\pi}) \; .$$

[14]For instance, to use the eight-point algorithm for 3D reconstruction.

References

[1] M. Born and E. Wolf. *Principles of Optics*. Pergamon Press, Oxford, 1975.

[2] R. P. Brent. An algorithm with guaranteed convergence for finding a zero of a function. In *Algorithms for Minimization without Derivatives*, pages 47–60. Prentice-Hall, Englewood Cliffs, NJ, 1973.

[3] Z. Zhang. Camera calibration. In G. Medioni and S. B. Kang, editors, *Emerging Topics in Computer Vision*, pages 4–43. Prentice-Hall, Englewood Cliffs, NJ, 2004.

Appendices

A Sensing

In a digital camera, still or video, the light that hits the image plane is collected by one or more *sensors*, that is, rectangular arrays of sensing elements. Each element is called a *pixel* (for "picture element"). The finite overall extent of the sensor array, together with the presence of diaphragms in the lens, limits the cone (or pyramid) of directions from which light can reach pixels on the sensor. This cone is called the *field of view* of the camera-lens combination.

Digital cameras have become pervasive in both the consumer and professional markets as well as in computer vision research. SLR (Single-Lens Reflex) still cameras are the somewhat bulkier cameras with an internal mirror that lets the photographer view the exact image that the sensor will see once the shutter button is pressed (hence the name: a single lens with a mirror (reflex)). These have larger sensors than CCTV cameras have, typically about 24 by 16 millimeters, although some very expensive models have sensors as large as 36 by 24 millimeters. More modern CCTV cameras are similar to the old ones, but produce a digital rather than analog signal directly. This signal is transferred to computer through a digital connection such as USB, or, for high-bandwidth video, IEEE 1394 (also known as Firewire), Apple Thunderbolt, or Gigabit Ethernet.

The next Section describes how pixels convert light intensities into voltages, and how these are in turn converted into numbers within the camera circuitry. This involves processes of integration (of light over the sensitive portion of each pixel), sampling (of the integral over time and at each pixel location), and addition of noise at all stages. These processes, as well as solutions for recording images in color, are then described in turn.

Pixels

A *pixel* on a digital camera sensor is a small rectangle that contains a photosensitive element and some circuitry. The photosensitive element is called a *photodetector*, or light detector. It is a semiconductor junction placed so that light from the camera lens can reach it. When a photon strikes the junction, it creates an electron-hole pair with approximately 70 percent probability (this probability is called the *quantum efficiency* of the detector). If the junction is part of a polarized electric circuit, the electron moves towards the positive pole and the hole moves towards the negative pole. This motion constitutes an electric current, which in turn causes an accumulation of charge (one electron) in a capacitor. A separate circuit discharges the capacitor at the beginning of the *shutter* (or *exposure*) interval. The charge accumulated over this interval of time is proportional to the amount of light that struck the capacitor during exposure, and therefore to the brightness of the part of the scene that the lens focuses on the pixel in question. Longer shutter times or greater image brightness both translate to more accumulated charge, until the capacitor fills up completely ("saturates").

> **Practical Aspects: CCD and CMOS Sensors.** Two methods are commonly used in digital cameras to read these capacitor charges: the CCD and the CMOS active sensor. The Charge-Coupled Device (CCD) is an electronic, analog shift register, and there is typically one shift register for each column of a CCD sensor. After the shutter interval has expired, the charges from all the pixels are transferred to the shift registers of their respective array columns. These registers in turn feed in parallel into a single CCD register at the bottom of the sensor, which transfers the charges out one row after the other as in a bucket brigade. The voltage across the output capacitor of this circuitry is proportional to the brightness of the corresponding pixel. A Digital to Analog (D/A) converter finally amplifies and transforms these voltages to binary numbers for transmission. In some cameras, the A/D conversion occurs on the camera itself. In

others, a separate circuitry (a frame grabber) is installed for this purpose on a computer that the camera is connected to.

The photodetector in a CMOS camera works in principle in the same way. However, the photosensitive junction is fabricated with the standard Complementary-symmetry Metal-Oxide-Semiconductor (CMOS) technology used to make common integrated circuits such as computer memory and processing units. Since photodetector and processing circuitry can be fabricated with the same process in CMOS sensors, the charge-to-voltage conversion that CCD cameras perform serially at the output of the CCD shift register can be done instead in parallel and locally at every pixel on a CMOS sensor. This is why CMOS arrays are also called Active Pixel Sensors (APS).

Because of inherent fabrication variations, the first CMOS sensors used to be much less consistent in their performance, both across different chips and from pixel to pixel on the same chip. This caused the voltage measured for a constant brightness to vary, thereby producing poor images at the output. However, CMOS sensor fabrication has improved dramatically in the recent past, and the two classes of sensors are now comparable to each other in terms of image quality. Although CCDs are still used where consistency of performance if of prime importance, CMOS sensors are eventually likely to supplant CCDs, both because of their lower cost and because of the opportunity to add more and more processing to individual pixels. For instance, "smart" CMOS pixels are being built that adapt their sensitivity to varying light conditions and do so differently in different parts of the image.

A Simple Sensor Model

Not all of the area dedicated to a pixel is necessarily photosensitive, as part of it is occupied by circuitry. The fraction of pixel area that collects light that can be converted to current is called the pixel's *fill factor*, and is expressed in percent. A 100 percent fill factor is achievable by covering each pixel with a properly shaped droplet of silica (glass) or silicon on each pixel. This droplet acts as a *micro-lens* that funnels photons from the entire pixel area onto the photo-detector. Not all cameras have micro-lenses, nor does a micro-lens necessarily work effectively on the entire pixel area. So different cameras can have very different fill factors. In the end, the voltage output from a pixel is the result of integrating light intensity over a pixel area determined by the fill factor.

The voltage produced is a nonlinear function of brightness. An approximate linearization is typically performed by a transformation called *gamma correction*,

$$V_{\text{out}} = V_{\max} \left(\frac{V_{\text{in}}}{V_{\max}} \right)^{1/\gamma}$$

where $V_{\max}$ is the maximum possible voltage and γ is a constant. Values of gamma vary, but are typically between 1.5 and 3, so V_{out} is a concave function of V_{in}, as shown in Figure 8: low input voltages are spread out at the expense of high voltages, thereby increasing the dynamic range[15] of the darker parts of the output image.

Noise affects all stages of the conversion of brightness values to numbers. First, a small current flows through the photodetectors even if no photons hit its junction. This source of imaging noise is called the *dark current* of the sensor. Typically, the dark current cannot be canceled away exactly, because it fluctuates somewhat and is therefore not entirely predictable. In addition, *thermal noise*, caused by the agitation of molecules in the various electronic devices and conductors, is added at all stages of the conversion, with or without light illuminating the sensor. This type of noise is well modeled by a Gaussian distribution. A third type of noise is the *shot noise* that is visible when the levels of exposure are extremely low (but nonzero).

[15] Dynamic range: in this context, this is the range of voltages available to express a given range of brightnesses.

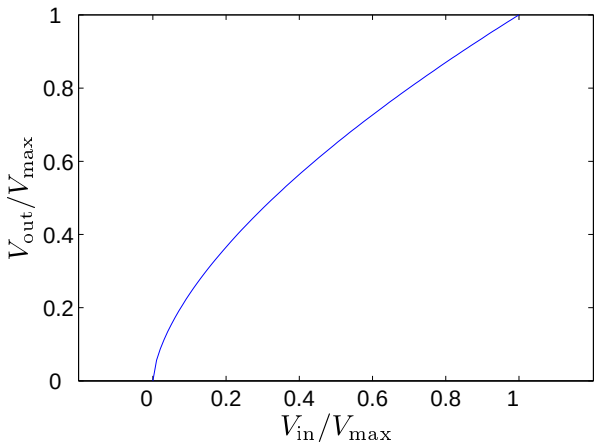


Figure 8: Plot of the normalized gamma correction curve for $\gamma = 1.6$.

In this situation, each pixel is typically hit by a very small number of photons within the exposure interval. The fluctuations in the number of photons are then best described by a Poisson distribution.

Every camera has *gain control* circuitry, either manually adjustable or automatic, which modifies the gain of the output amplifier so that the numerical pixel values occupy as much of the available range as possible. With dark images, the gain is set to a large value, and to a small value for bright ones. Gain is typically expressed in ISO values, from the standard that the International Standardization Organization (ISO) has defined for older film cameras. The ISO scale is linear, in the sense that doubling the ISO number corresponds to doubling the gain.

If lighting in the scene cannot be adjusted, a dark image can be made brighter by either (i) opening the lens aperture or (ii) by increasing exposure time, or (iii) by increasing the gain. The effects, however, are very different. As discussed earlier, widening the aperture decreases the depth of field. Increasing exposure time may result into blurry images if there is motion in the scene.

Figure 9 shows the effect of different gains. The two pictures were taken with constant lighting and aperture. However, the one in (a) (and the detail in (c)) was taken with a low value of gain, and the one in (b) (and (d)) was taken with a gain value sixteen times greater. From the image as a whole ((a) and (b)) one can notice some greater degree of "graininess" corresponding to a higher gain value. The difference is more obvious when details of the images are examined ((c) and (d)).

So there is no free lunch: more light is better for a brighter picture. That is, brightness should be achieved by shining more light on the scene or, if depth of field is not important, by opening the aperture. Increasing camera gain will make the picture brighter, but also noisier.

In summary, a digital sensor can be modeled as a light integrator over an area corresponding to the pixel's fill factor. This array is followed by a sampler, which records the values of the integral at the centers of the pixels. At the output, an adder adds noise, which is an appropriate combination of dark current, Gaussian thermal noise, and shot noise. The parameters of the noise distribution typically depend on brightness values and camera settings. Finally, a quantizer converts continuous voltage values into discrete pixel values. The gamma correction can be ignored if the photodetectors are assumed to have an approximately linear response. Figure 10 shows this model in diagram form.

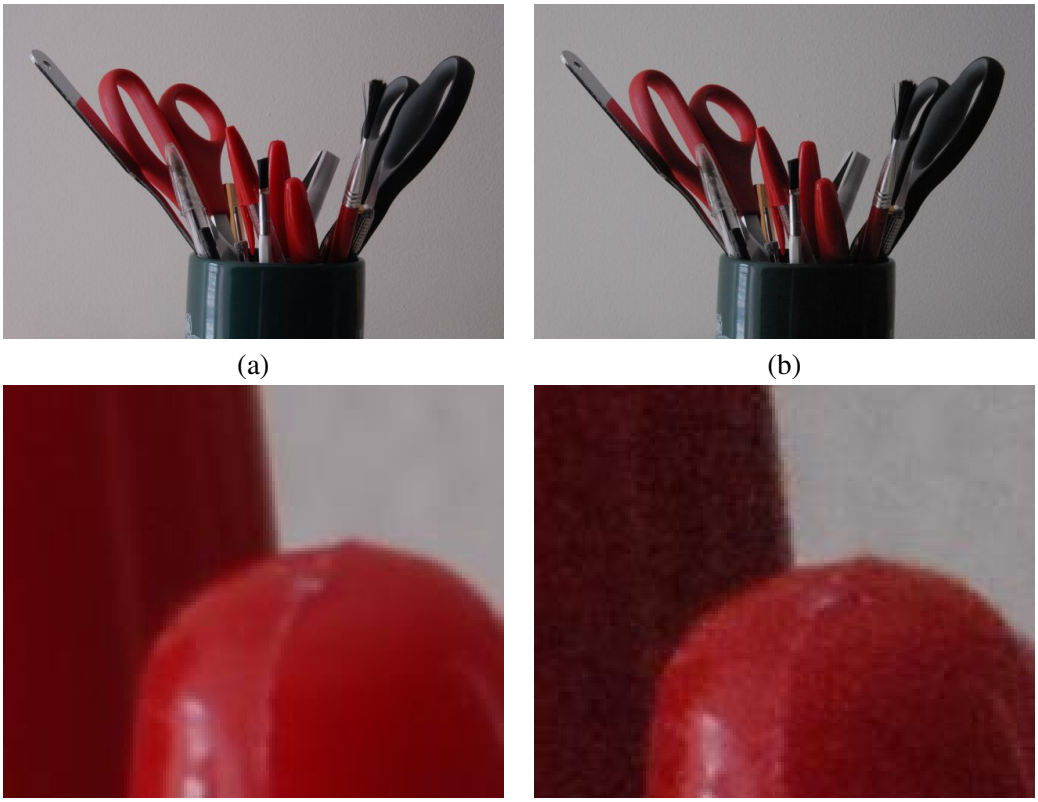


Figure 9: These two images were taken with the same lens aperture of f/20. However, (a) was taken with a low gain setting, corresponding to sensitivity ISO 100, and a one-second exposure, while (b) was taken with a high gain setting of ISO 1600, and an exposure of 1/15 of a second. (c) and (d) show the same detail from (a) and (b), respectively.

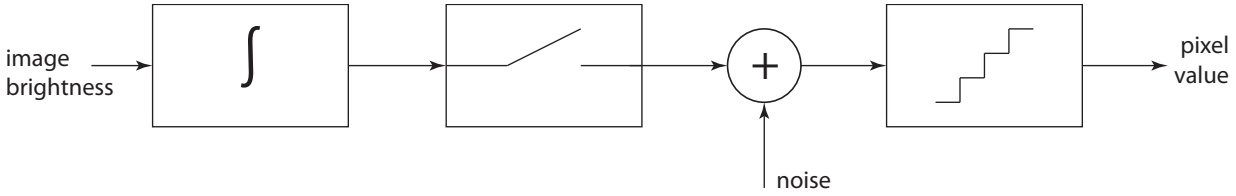


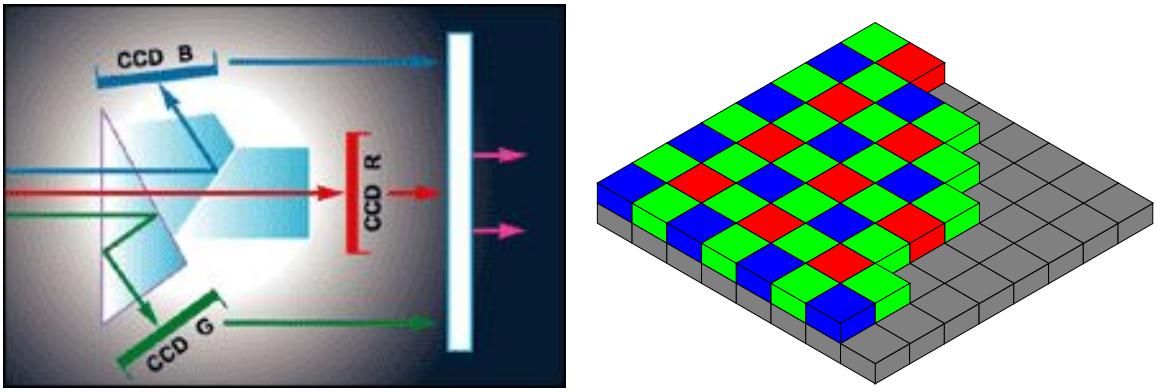
Figure 10: A simple sensor model. The three rectangular boxes are an integrator, a sampler, and a quantizer. Both integrator and samplers are in two dimensions. Noise statistics depend on input brightness and on camera settings.

Color Sensors

The photodetectors in a camera sensor are only sensitive to light brightness, and do not report color. Two standard methods are used to obtain color images. The first, the 3-sensor method, is expensive and high quality. The second, the Bayer mosaic, is less expensive and sacrifices resolution for color. These two methods are discussed in turn.

The 3-Sensor Method In a 3-sensor color camera, a set of glass prisms uses a combination of internal reflection and refraction to split the incoming image into three. The three beams exit from three different faces of the prism, to which three different sensor arrays are attached. Each sensor is coated with a die that lets only light in a relatively narrow band go through in the red, green, and blue parts of the spectrum, respectively. Figure 11 (a) shows a schematic diagram of a beam splitter.

The Bayer Mosaic A more common approach to color imaging is the *sensor mosaic*. This scheme uses a single sensor, but colors the micro-lenses of individual pixels with red, green, or blue die. The most common pattern is the so-called *Bayer mosaic*, shown in Figure 11 (b).



(a) (b)

Figure 11: (a) Schematic diagram of a 3-sensor beam splitter for color cameras. From `http://www.usa.canon.com/`. (b) The Bayer color pattern. From `http://en.wikipedia.org/wiki/Image:Bayer_pattern_on_sensor.svg`.

With this arrangement, half of the pixels are sensitive to the green band of the light spectrum, and one quarter each to blue and red. This is consistent with the distribution of color-sensitive cones in the human retina, which is more responsive to the green-yellow part of the spectrum than to its red or blue components.

The raw image produced with a Bayer mosaic contains one third of the information that would be obtained with a 3-sensor camera of equal resolution on each chip. While each point in the field of view is seen by three pixels in a 3-sensor camera, *no* point in the world is seen by more than one pixel in the Bayer mosaic. As a consequence, the blue and red components of a pixel that is sensitive only to the green band must be inferred, and an analogous statement holds for the other two types of pixels. After properly normalizing and gamma-correcting each pixel value, this inference proceeds by interpolation, under the assumption that nearby pixels usually have similar colors.

Practical Aspects: 3-Sensor Versus Bayer. Of course, the beam splitter and the additional two sensors add cost to a 3-sensor camera. In addition, the three sensors must be aligned very precisely on the faces

of the beam splitter. This fabrication aspect has perhaps an even greater impact on final price than the trebled sensor cost. Interestingly, even high-end SLR cameras use the Bayer mosaic for color, as the loss of information caused by mosaicing is usually satisfactorily compensated by sensor resolutions in the tens of millions of pixels.

B Rotation Vectors

In numerical optimization problems, the redundancy of 3×3 rotation matrices is inconvenient, and a minimal, three-parameter representation of rotation is often preferable.

The simplest such representation is based on *Euler's theorem*, stating that every rotation can be described by an axis of rotation and an angle around it. A compact representation of axis and angle is a three-dimensional *rotation vector* whose direction is the axis and whose magnitude is the angle in radians. The axis is oriented so that the rotation with the smaller angle is counterclockwise around it. As a consequence, the angle of rotation is always nonnegative, and at most π.

While simple, the rotation-vector representation of rotation must be used with some care. As defined above, the set of all rotation vectors is the three-dimensional ball[16] of radius π. However, while points in the interior of the ball represent distinct rotations, two antipodal points on its surface, that is, two vectors $\mathbf{r}$ and $-\mathbf{r}$ with norm π, represent the same 180-degree rotation.

Whether this lack of uniqueness is a problem depends on the application. For instance, when comparing rotations, it would be troublesome if the same rotation had two different representations. To preserve uniqueness, one can carefully peel away half of the sphere from the ball, and define the *half-open rotation ball* as the following union of disjoint sets:

$$\{\mathbf{r} \,:\, \|\mathbf{r}\| < \pi\} \,\cup\, \{\mathbf{r} \,:\, \|\mathbf{r}\| = \pi \cap r_1 > 0\} \,\cup\, \{\mathbf{r} \,:\, \|\mathbf{r}\| = \pi \cap r_1 = 0 \cap r_2 > 0\} \,\cup\, \{(0, 0, \pi)\} \,.$$

These sets are respectively the open unit ball, the open hemisphere with its pole at $(\pi, 0, 0)$, the open half-equator of that hemisphere centered at $(0, \pi, 0)$, and the individual point $(0, 0, \pi)$. The last three sets are illustrated in Figure 12.

The formula for finding the rotation matrix corresponding to an angle-axis vector is called *Rodrigues' formula*, which is now derived.

Let $\mathbf{r}$ be a rotation vector. If the vector is $(0, 0, 0)$, then the rotation is zero, and the corresponding matrix is the identity matrix:

$$\mathbf{r} = \mathbf{0} \rightarrow R = I \,.$$

Let us now assume that $\mathbf{r}$ is not the zero vector. The unit vector for the axis of rotation is then

$$\mathbf{u} = \frac{\mathbf{r}}{\|\mathbf{r}\|}$$

and the angle is

$$\theta = \|\mathbf{r}\| \quad \text{radians.}$$

The rotation has no effect on a point $\mathbf{p}$ along the axis. Suppose then that $\mathbf{p}$ is off the axis. To see the effect of rotation on $\mathbf{p}$, we decompose $\mathbf{p}$ into two orthogonal vectors, one along $\mathbf{u}$ and the other perpendicular to it:

$$\mathbf{a} = P_{\mathbf{u}}\mathbf{p} = \mathbf{u}\mathbf{u}^T\mathbf{p}$$

[16] A *ball* of radius r in $\mathbb{R}^n$ is the set of points $\mathbf{p}$ such that $\|\mathbf{p}\| \leq r$. In contrast, a *sphere* of radius r in $\mathbb{R}^n$ is the set of points $\mathbf{p}$ such that $\|\mathbf{p}\| = r$.

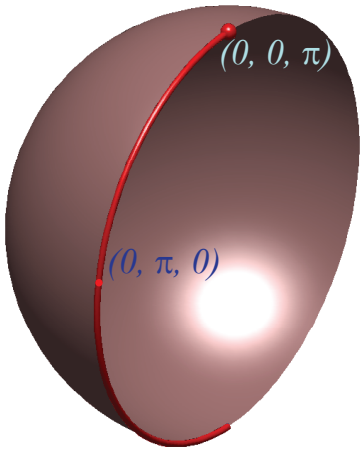


Figure 12: The parts of the sphere of radius π that are included in the half-open rotation ball. The interior of the ball is included as well, but is not shown in this figure for clarity. The pole of the hemisphere in the picture is the point $(\pi, 0, 0)$.

is along $\mathbf{u}$, and

$$\mathbf{b} = \mathbf{p} - \mathbf{a} = (1 - \mathbf{u}\mathbf{u}^T)\mathbf{p}$$

is orthogonal to $\mathbf{u}$, as shown in Figure 13.

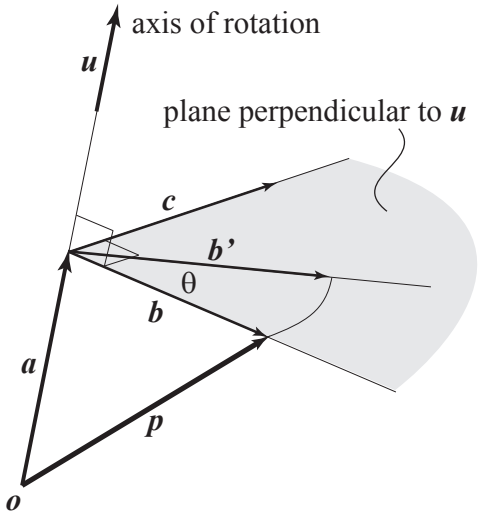


Figure 13: Vectors used in the derivation of Rodrigues' formula.

The rotation leaves $\mathbf{a}$ unaltered, and rotates $\mathbf{b}$ by θ in the plane orthogonal to $\mathbf{u}$. To express the latter rotation, we introduce a third vector

$$\mathbf{c} = \mathbf{u} \times \mathbf{p}$$

that is orthogonal to both $\mathbf{u}$ and $\mathbf{p}$, and has the same norm as $\mathbf{b}$ (because $\mathbf{u}$ is a unit vector). Since $\mathbf{b}$ and $\mathbf{c}$ have the same norm, the rotated version of $\mathbf{b}$ is

$$\mathbf{b}' = \mathbf{b}\cos\theta + \mathbf{c}\sin\theta \ .$$

The rotated version of the entire vector $\mathbf{p}$ is then

$$\begin{aligned}\mathbf{p}' &= \mathbf{a} + \mathbf{b}' = \mathbf{a} + \mathbf{b}\cos\theta + \mathbf{c}\sin\theta = \mathbf{u}\mathbf{u}^T\mathbf{p} + (1 - \mathbf{u}\mathbf{u}^T)\mathbf{p}\cos\theta + \mathbf{u}\times\mathbf{p}\sin\theta \\ &= [I\cos\theta + (1-\cos\theta)\mathbf{u}\mathbf{u}^T + \mathbf{u}_\times\sin\theta]\mathbf{p}\end{aligned}$$

so that

$$R = I\cos\theta + (1-\cos\theta)\mathbf{u}\mathbf{u}^T + \mathbf{u}_\times\sin\theta\ .$$

This equation is called *Rodrigues' formula*.

To invert this formula, note that the sum of its first two terms,

$$I\cos\theta + (1-\cos\theta)\mathbf{u}\mathbf{u}^T$$

is a symmetric matrix, while the last term,

$$\mathbf{u}_\times\sin\theta$$

is antisymmetric. Therefore,

$$R - R^T = 2\mathbf{u}_\times\sin\theta = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}\sin\theta = 2\begin{bmatrix} 0 & -\rho_3 & \rho_2 \\ \rho_3 & 0 & -\rho_1 \\ -\rho_2 & \rho_1 & 0 \end{bmatrix}\ .$$

Since the vector $\mathbf{u}$ has unit norm, the norm of the vector (ρ_1, ρ_2, ρ_3) is $\sin\theta$. Direct calculation shows that the *trace*, that is, the sum of the diagonal elements of the rotation matrix R, is equal to $2\cos\theta + 1$, so that

$$\cos\theta = (r_{11} + r_{22} + r_{33} - 1)/2\ .$$

If $\sin\theta = 0$ and $\cos\theta = 1$ then the rotation vector is

$$\mathbf{r} = \mathbf{0}\ .$$

If $\sin\theta = 0$ and $\cos\theta = -1$ then Rodrigues' formula simplifies to the following:

$$R = -I + 2\mathbf{u}\mathbf{u}^T$$

so that

$$\mathbf{u}\mathbf{u}^T = \frac{R + I}{2}\ .$$

This equation shows that each of the three columns of $(R + I)/2$ is a multiple of the unknown unit vector $\mathbf{u}$. Since the norm of $\mathbf{u}$ is one, not all its entries can be zero. Let $\mathbf{v}$ be any nonzero column of $R + I$. Then

$$\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

and

$$\mathbf{r} = \mathbf{u}\pi\ .$$

Finally, in the general case, $\sin\theta \neq 0$. Then, the normalized rotation vector is

$$\mathbf{u} = \frac{\rho}{\|\rho\|}\ .$$

From $\sin\theta$ and $\cos\theta$, the two-argument arc-tangent function yields the angle θ, and

$$\mathbf{r} = \mathbf{u}\theta \ .$$

Recall that the two-argument function $\arctan_2$ is defined as follows for $(x, y) \neq (0, 0)$:

$$\arctan_2(y, x) = \begin{cases} \arctan(\frac{y}{x}) & \text{if } x > 0 \\ \pi + \arctan(\frac{y}{x}) & \text{if } x < 0 \\ \frac{\pi}{2} & \text{if } x = 0 \text{ and } y > 0 \\ -\frac{\pi}{2} & \text{if } x = 0 \text{ and } y < 0 \end{cases}$$

and is undefined for $(x, y) = (0, 0)$. This function returns the arc-tangent of y/x (notice the order of the arguments) in the proper quadrant, and extends the function by continuity along the y axis.

Table 1 summarizes this discussion.

C Initialization of Calibration Parameters

This appendix shows how to find an initial, approximate set of camera calibration parameters by setting distortion to zero. In the absence of distortion, the camera model becomes

$$\begin{aligned} \mathbf{X} &= R\,(\mathbf{W} - \mathbf{t}) \\ \mathbf{x} &= p(\mathbf{X}) \ = \ \frac{1}{X_3} \left[\begin{array}{c} X_1 \\ X_2 \end{array} \right] \\ \boldsymbol{\xi} &= S\mathbf{x} + \boldsymbol{\pi} \ , \end{aligned}$$

and it is easy to verify that the second and third equation can be repackaged into the following equation:

$$X_3 \boldsymbol{\xi} = A\mathbf{X} \quad \text{where} \quad A = \left[\begin{array}{c|c} S & \boldsymbol{\pi} \end{array} \right] \tag{5}$$

is a 2×3 matrix. Then the right-hand side $A\mathbf{X}$ of the equation above can be written as follows:

$$A\mathbf{X} = AR\,(\mathbf{W} - \mathbf{t}) = \tilde{B}\mathbf{W} - \tilde{B}\mathbf{t} \quad \text{where} \quad \tilde{B} = AR \ . \tag{6}$$

For convenience in the manipulations that follow, we define the vector

$$\mathbf{a} = \left[\begin{array}{c} a_1 \\ a_2 \\ a_3 \end{array} \right] = -B\mathbf{t} \quad \text{where} \quad B = \left[\begin{array}{c} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \mathbf{b}_3^T \end{array} \right] = \left[\begin{array}{c} \tilde{B} \\ \mathbf{k}^T \end{array} \right] \tag{7}$$

and where $\mathbf{b}_3^T = \mathbf{k}^T$ is the third row of R:

$$R = \left[\begin{array}{c} \mathbf{i}^T \\ \mathbf{j}^T \\ \mathbf{k}^T \end{array} \right] \ .$$

Thus, since

$$X_3 = \mathbf{k}^T(\mathbf{W} - \mathbf{t}) = \mathbf{k}^T\mathbf{W} + a_3 \ ,$$

we can combine equations (5) and (6) and replace $\mathbf{k}$ with $\mathbf{b}_3$ to yield the following system of two linear equations in B and $\mathbf{a}$:

$$\begin{aligned} \mathbf{b}_1^T\mathbf{W} + a_1 - (\mathbf{b}_3^T\mathbf{W} + a_3)\,\xi_1 &= 0 \\ \mathbf{b}_2^T\mathbf{W} + a_2 - (\mathbf{b}_3^T\mathbf{W} + a_3)\,\xi_2 &= 0\,. \end{aligned}$$

One such system of two equations can be written for each world point $\mathbf{W}_n$ and corresponding image point $\boldsymbol{\xi}_n$. To this end, we first introduce the 4-dimensional vector

$$\mathbf{w}_n = \begin{bmatrix} \mathbf{W}_n \\ 1 \end{bmatrix}$$

and write the system above in matrix form as

$$Q_n\mathbf{q} = \mathbf{0}_2 \quad \text{where} \quad Q_n = \begin{bmatrix} \mathbf{w}_n^T & \mathbf{0}_4^T & -\xi_{n1}\mathbf{w}_n^T \\ \mathbf{0}_4^T & \mathbf{w}_n^T & -\xi_{n2}\mathbf{w}_n^T \end{bmatrix} \quad \text{and} \quad \mathbf{q} = \begin{bmatrix} \mathbf{b}_1 \\ a_1 \\ \mathbf{b}_2 \\ a_2 \\ \mathbf{b}_3 \\ a_3 \end{bmatrix}$$

and where $\mathbf{0}_k$ is a column vector of k zeros. The matrix Q_n is 2×12 and the vector $\mathbf{q}$ is 12×1. If N points are available, one can build the $2N \times 12$ system

$$Q\mathbf{q} = \mathbf{0}_{2N} \quad \text{where} \quad Q = \begin{bmatrix} Q_1 \\ \vdots \\ Q_N \end{bmatrix}$$

whose least-squares solution yields the unknowns in $\mathbf{q}$. This solution is unique if the last singular value of Q is strictly smaller than all the others, which requires $N \geq 6$ (a necessary but not sufficient condition). Since the system is homogenous, the solution is defined up to a scaling factor. However, we know that $\mathbf{q}(9:11) = \mathbf{b}_3 = \mathbf{k}$, a unit vector, so that we normalize $\mathbf{q}$ as follows:

$$\mathbf{q} \leftarrow \frac{\mathbf{q}}{\|\mathbf{q}(9:11)\|}\,.$$

To recover the camera parameters R, $\mathbf{t}$, $\boldsymbol{\pi}$, s_1, s_2 from the matrix B and vector $\mathbf{a}$ extracted from the normalized solution $\mathbf{q}$, we note that

$$\tilde{B} = B(1{:}2,:) \quad \text{and} \quad \mathbf{k}^T = B(3,:)$$

so that we can compute

$$\tilde{B}\mathbf{k} = AR\mathbf{k} = A\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{\pi}$$

and

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \tilde{B}\tilde{B}^T = ARR^TA^T = AA^T = \begin{bmatrix} s_1^2 + \pi_1^2 & \pi_1\pi_2 \\ \pi_1\pi_2 & s_2^2 + \pi_2^2 \end{bmatrix}\,.$$

The two diagonal entries of C yield the two entries on the diagonal of the diagonal matrix S:

$$s_1 = \sqrt{c_{11} - \pi_1^2} \quad \text{and} \quad s_2 = \sqrt{c_{22} - \pi_2^2}\,.$$

Knowing $\boldsymbol{\pi}$ and S yields A from its definition in equation (5). Finally, R and $\mathbf{t}$ can be found by solving the two linear systems

$$\begin{bmatrix} & A & \\ 0 & 0 & 1 \end{bmatrix} R = B \quad \text{and} \quad -B\mathbf{t} = \mathbf{a} \; .$$

The third row in the first equation above is added to ensure that the third row of R to is equal to the third row of B, that is, to $\mathbf{k}^T$. The first two rows are from equation (6). The second equation above is the definition of $\mathbf{a}$ given in equation (7).

If the determinant of the resulting matrix R is negative, the whole unpacking procedure can be repeated starting with $-\mathbf{q}$ rather than $\mathbf{q}$. Finally, since R results from solving a homogeneous system that does not impose any constraint on the solution (other than unit-norm $\mathbf{q}$), R is not necessarily orthogonal. The orthogonal matrix that is closest to R can be computed by first taking the SVD of R,

$$R = U\Sigma V^T$$

and then replacing R by UV^T.

The rotation matrix R corresponding to the rotation vector $\mathbf{r}$ such that $\|\mathbf{r}\| \leq \pi$ can be computed as follows. Let

$$\theta = \|\mathbf{r}\|$$

If $\theta = 0$, then $R = I$. Otherwise,

$$\mathbf{u} = \frac{\mathbf{r}}{\theta} \quad \text{and} \quad R = I\cos\theta + (1 - \cos\theta)\mathbf{u}\mathbf{u}^T + \mathbf{u}_\times \sin\theta \ .$$

Conversely, the rotation vector corresponding to the rotation matrix

$$R \quad \text{such that} \quad R^T R = RR^T = I \quad \text{and} \quad \det(R) = 1$$

can be computed as follows. Let

$$A = \frac{R - R^T}{2} \quad , \quad \rho = \begin{bmatrix} a_{32} & a_{13} & a_{21} \end{bmatrix}^T \quad , \quad s = \|\rho\| \quad , \quad c = (r_{11} + r_{22} + r_{33} - 1)/2 \ .$$

If $s = 0$ and $c = 1$, then $\mathbf{r} = \mathbf{0}$. Otherwise, if $s = 0$ and $c = -1$, let $\mathbf{v}$ = a nonzero column of $R + I$. Then,

$$\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|} \quad , \quad \mathbf{r} = S_{1/2}(\mathbf{u}\pi) \ .$$

Finally, if $\sin\theta \neq 0$,

$$\mathbf{u} = \frac{\rho}{s} \quad , \quad \theta = \arctan_2(s, c) \quad \text{, and} \quad \mathbf{r} = \mathbf{u}\theta \ .$$

The function $S_{1/2}(\mathbf{r})$ flips signs of the coordinates of vector $\mathbf{r}$ (assumed here to have norm π) to force it onto the half-hemisphere of Figure 12, in order to ensure uniqueness:

$$S_{1/2}(\mathbf{r}) = \begin{cases} -\mathbf{r} & \text{if } \|\mathbf{r}\| = \pi \text{ and } ((r_1 = r_2 = 0 \text{ and } r_3 < 0) \\ & \quad \text{or } (r_1 = 0 \text{ and } r_2 < 0) \text{ or } (r_1 < 0)) \\ \mathbf{r} & \text{otherwise.} \end{cases}$$

Table 1: Transformations between a rotation matrix R and a rotation vector $\mathbf{r}$.