Perspective Projection
Problem Set 2 Solutions

Computer Vision 2021
University of Bern

1 Camera models

1. **Thin lens formula**

   A camera is equipped with a lens with focal length 50mm and aperture diameter $D = 2$cm. An object at 5m from the camera is in focus. What is the minimum thickness of the camera?

   Solution The thickness of the camera is lower bounded by the distance image plane to lens plane. According to the think lens law we know that

   $$\frac{1}{5} + \frac{1}{v} = \frac{1}{50 \cdot 10^{-3}} \tag{1}$$

   where v is the image plane to lens plane distance. Then, we have $v = 5.05$cm.

2. **Thin lens formula**

   A camera is equipped with a lens with focal length 40mm and aperture diameter $D = 2$cm. The distance of the image plane to the lens plane is 5cm. What is the blur diameter of an object at infinity?

   Solution According to the thin lens law we know that

   $$\frac{1}{\infty} + \frac{1}{u} = \frac{1}{40 \cdot 10^{-3}} \tag{2}$$

   where u is distance from the lens (inside the camera) where the object at infinity is in focus. By similar triangles, we have that the blur diameter satisfies

$$\frac{D}{u} = \frac{B}{|5 \cdot 10^{-2} - u|} \tag{3}$$

where B is the blur diameter. By putting all together we find $u = 40 \cdot 10^{-3}$ and $B = 0.5$cm.

3. **Field of view**

   A simple camera is made of a CCD sensor and a single lens. The size of a CCD sensor is 4cm $\times$ 4cm. The focal length of the lens is $f = 50$mm. What is the field of view of this camera?

   Solution The field of view can be computed according to

   $$\phi = 2 \arctan \frac{S}{2f} = 2 \arctan \frac{4 \cdot 10^{-2}}{2 \cdot 50 \cdot 10^{-3}} = 43.6^\circ, \tag{4}$$

   where S is the size of the sensor.

4. **Geometric optics**

   A simpler way to handle lenses and to determine the imaging equations is to use a *ray representation*. Instead of considering the 3D point in space, we consider a ray in space (a line) and how this is deflected by a lens when it falls within its aperture. Let us consider a lens with focal length f and a line incident at an angle θ with the lens (the angle is with respect to the normal to the lens plane) at a (signed) height d from the optical center. Compute the exiting angle ϕ by using the two basic rules of a thin lens: 1) rays through the optical center are not deflected and 2) objects at infinity are focused at a distance f behind the lens (this means that parallel rays converge at a point at a distance f behind the lens). See Fig. 1.

   Solution If we take the ray with the same incident angle θ through the optical center, we know that it will not be deflected behind the lens.

   Also, the deflected ray from the incident one at a (signed) height d will intersect the previous one at a distance f. In equations, we have

   $$f = \rho \cos \theta \tag{5}$$
   $$a = \rho \sin \theta = f \tan \theta \tag{6}$$

   where ρ is the length of the segment behind the lens from the optical center to the point in focus and a is the (signed) height (on the lens plane) of the point in focus from the optical center.

   $$d - a = -b \sin \phi \tag{7}$$
   $$f = -b \cos \phi = \frac{a - d}{\tan \phi} \tag{8}$$

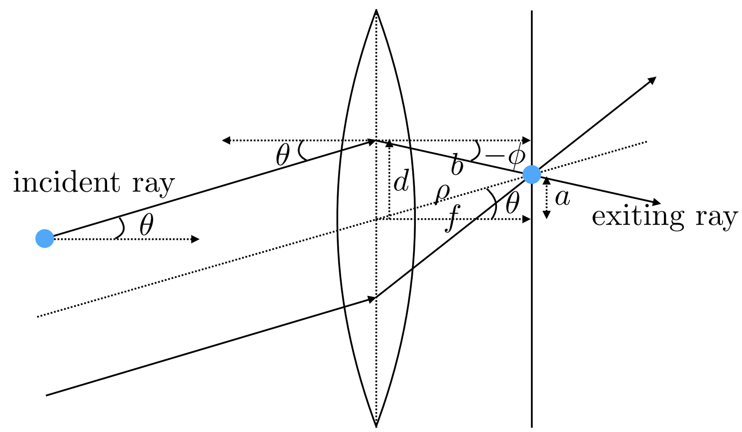

Fig. 1: Geometric optics.

where b is the length of the segment of the deflected ray behind the lens up to the point in focus. So now we obtain

$$\tan\phi = \frac{f\tan\theta - d}{f} = \tan\theta - \frac{d}{f}. \tag{9}$$

The equation above is the update to the direction of a ray after going through the aperture of a lens.

Let us test the equation with a few cases. If $\theta = 0$ the rays are all parallel to the optical axis. In this case $\tan\phi = -\frac{d}{f}$. Simple geometric verification shows that this result is correct. Another example is when $d = 0$, *i.e.*, when the ray goes through the optical center and it is not deflected. Then $\theta = \phi$ which is also correct in our notation. Another test is that ϕ increases as d decreases or as θ increases (within the range $[-\pi/2, \pi/2]$).

To make our geometric optics calculations complete, let us represent a ray with the 3D vector $[\tan\theta\ d\ 1]^T$ where θ is the incident angle and d the height (with respect to the optical axis). Then, we already know how to update the 3D representation when it meets a lens with focal length f. We can use the matrix

$$M_L = \begin{bmatrix} 1 & -\frac{1}{f} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{10}$$

So that we have

$$V_{ext} = M_L V_{inc} = \begin{bmatrix} 1 & -\frac{1}{f} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tan\theta \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} \tan\theta - \frac{d}{f} \\ d \\ 1 \end{bmatrix} \tag{11}$$

If the lens is not centered at the optical axis, but its center has height b from the optical axis, we can easily generalize the matrix M_L to

$$M_L = \begin{bmatrix} 1 & -\frac{1}{f} & \frac{b}{f} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

as what matters in the lens calculation is the height from the lens center $(d - b)$.

We also consider empty space as an optical element. If we have an empty space of Δ, we can update the ray with this matrix

$$M_S = \begin{bmatrix} 1 & 0 & 0 \\ \Delta & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{13}$$

so that we obtain

$$V_{ext} = M_S V_{inc} = \begin{bmatrix} 1 & 0 & 0 \\ \Delta & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tan\theta \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} \tan\theta \\ d + \Delta \tan\theta \\ 1 \end{bmatrix}. \tag{14}$$

A ray moving in space does not deflect, but its height changes.

With these few building blocks we can perform approximate *ray tracing*, that is we can simulate the path of a ray as it goes through a combination of lenses and space.

5. **Composite optical systems (challenging)**

   Let us consider a camera with 2 lenses and let us use the geometric optics approximation to determine the imaging equations. Suppose that the first lens is centered at the optical axis and has focal length f_1. The second lens is behind the first lens at a distance Δ_1, has focal length f_2 and its center is displaced at a height b from the optical axis. For simplicity we assume that both lenses have infinite aperture. The camera sensor is behind the second lens at a distance Δ_2. A point P is in front of the lens and emits light rays. See Fig. 2.

   Determine the ray tracing matrix M of the whole imaging system (to determine the propagation of the rays from P).

Solution The composite matrix M is obtained by chain multiplication. First we have the space between a point in space and the first lens M_{S0}. Then, the first lens M_{L1}, the second space M_{S1}, the second lens M_{L2}, and finally the space between the second lens and the image plane M_{S2}. Thus,

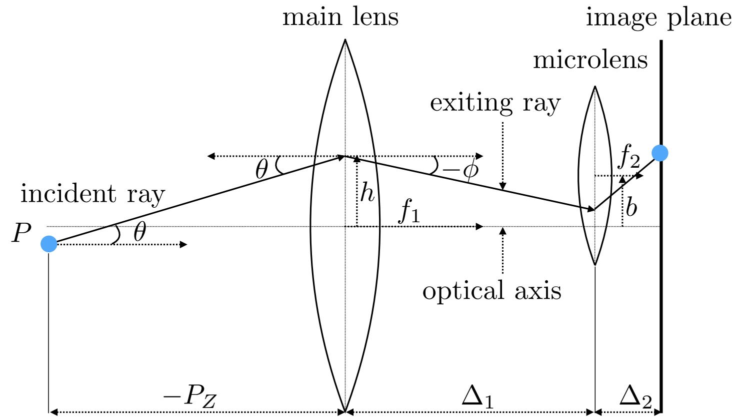

Fig. 2: Geometric optics with multiple lenses.

we have

$$\begin{aligned}
M &= M_{S2}M_{L2}M_{S1}M_{L1}M_{S0} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ \Delta_2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{f_2} & \frac{b}{f_2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ \Delta_1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{f_1} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ P_Z & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (15) \\
&= \begin{bmatrix} 1 & -\frac{\Delta_2}{f_2} & \frac{b}{f_2} \\ \Delta_2 & 1-\frac{\Delta_2}{f_2} & \frac{b\Delta_2}{f_2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{\Delta_1}{f_1} & 0 \\ \Delta_1 & 1-\frac{\Delta_1}{f_1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ P_Z & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1-\frac{\Delta_1\Delta_2}{f_2} & -\frac{\Delta_1}{f_1}-\frac{\Delta_2}{f_2}\left(1-\frac{\Delta_1}{f_1}\right) & \frac{b}{f_2} \\ \Delta_2+\Delta_1\left(1-\frac{\Delta_2}{f_2}\right) & -\frac{\Delta_1\Delta_2}{f_1}+\left(1-\frac{\Delta_2}{f_2}\right)\left(1-\frac{\Delta_1}{f_1}\right) & \frac{b\Delta_2}{f_2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ P_Z & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.
\end{aligned}$$

It is easy to see how the complexity of the imaging matrix quickly grows as we compose different optical elements.

2 Filters

1. **Median filter**

   The median filter takes the middle element after sorting the elements in a patch. While this operation is quite robust to non Gaussian noise, it can lead to severe flattening of the texture when the size of the patch is large. How could we modify the median filter so that it retains the original texture while removing outliers?

Solution One possible modification is to leave the center pixel if this is within a certain distance from the median pixel of the patch (*e.g.*, below the first quartile or above the third quartile).

2. **Nonlocal means filter (challenging)**

The nonlocal means filter performs averaging of patches that are sufficiently similar. If we extract a patch P from the noisy image, then we select other patches P_i such that

$$|P - P_i| < d \tag{16}$$

with a certain scalar $d > 0$. If N patches P_i satisfy the constraint above, then the updated/denoised patch $\hat{P}$ is obtained via averaging

$$\hat{P} = \frac{1}{N}\sum_{i=1}^{N} P_i. \tag{17}$$

Suppose that the original image is affected by zero-mean i.i.d. Gaussian noise. Then, we can write

$$P = P_0 + n \qquad\qquad P_i = P_0 + n_i \tag{18}$$

where P_0 is the noise-free patch and $n, n_i \sim \mathcal{N}(0, \sigma^2 I_d)$ are Gaussian random variables. What is the distribution of $\hat{P}$?

Solution The average of Gaussian random variables is also a Gaussian random variable. Its mean is the average of the means, which are all zero in this case. The variance is the average of the individual variances (which are all identical)

$$\tilde{\sigma}^2 = \frac{1}{N^2}(N\sigma^2) = \frac{1}{N}\sigma^2. \tag{19}$$

The distribution of the average of the patches tends to peek around P_0, which is the noise-free patch.

3. **Mean shift filter (challenging)**

The *mean shift* filter is a procedure to find the maxima of a probability density function (the modes) from its samples. The first step is to assume that the probability density function can be approximated by its kernel density estimate

$$p(x) = \frac{1}{N}\sum_{i=1}^{N} K\left(\frac{|x - x_i|^2}{h}\right) \tag{20}$$

where $K(|x|^2)$ is a kernel that integrates to 1 in x and h is a bandwidth parameter. A necessary condition to determine the maxima of a function is that the first order derivatives must be zero. By taking the first order derivatives in x we obtain

$$\nabla p(x) = \frac{1}{N}\sum_{i=1}^{N} 2\frac{x - x_i}{h} K'\left(\frac{|x - x_i|^2}{h}\right) = 0 \tag{21}$$

Next we rearrange the equation above to obtain the following iterative update

$$x^{t+1} = \frac{\sum_{i=1}^{N} x_i K'\left(\frac{|x^t - x_i|^2}{h}\right)}{\sum_{i=1}^{N} K'\left(\frac{|x^t - x_i|^2}{h}\right)} \tag{22}$$

This update is the *mean shift* algorithm and it can be shown to converge to the modes of the distribution $p(x)$ in a number of cases. Let us consider the Epanechnikov kernel

$$K(z) = \begin{cases} \frac{3}{4}(1 - z) & 0 \leq z \leq 1 \\ 0 & z > 1 \end{cases} \tag{23}$$

then, we have

$$K'(z) = \begin{cases} -\frac{3}{4} & 0 \leq z \leq 1 \\ 0 & z > 1 \end{cases} \tag{24}$$

and the meanshift iteration becomes

$$x^{t+1} = \frac{\sum_{i=1}^{N} x_i \delta(|x^t - x_i|^2 < h)}{\sum_{i=1}^{N} \delta(|x^t - x_i|^2 < h)} \tag{25}$$

where $\delta(\text{event})$ is the indicator function of "event" (it is 1 if the event is true and 0 otherwise). Find the relation between the mean shift filter with Epanechnikov kernel and the nonlocal mean filter above.

Solution The mean shift filter is the repeated application of the nonlocal mean filter.

3 Edges

1. **Gradients**

   Let us consider an image x. Let also $p, q \in \mathbf{R}^2$ be two pixels (represented as two 2D vectors) in x. We define *self-similarity* as the property that

$$x[q] = x[\alpha(q - p) + p] \qquad \forall q : |q - p| \leq \rho \tag{26}$$

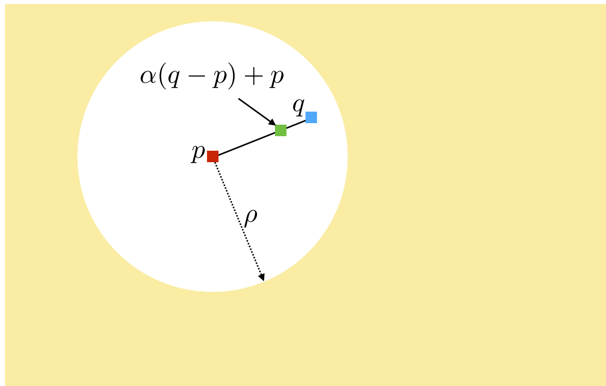

Fig. 3: Gradients and self-similarity.

where $\rho > 0$ is the radius of a ball around p and $0 < \alpha < 1$. See Fig. 3. In other words, rays originating at p should have constant image intensity. Notice that, in particular, the self-similarity property is satisfied at corners and at edges. How does the gradient $\nabla x[q]$ in the point q relate to the self-similarity in the limit as $\alpha \to 1$?

Solution The self-similarity property is equivalent to

$$x[q] - x[\alpha(q-p)+p] = 0 \qquad \forall q : |q-p| \leq \rho \tag{27}$$

Then, we obtain

$$\lim_{\alpha \to 1} \frac{x[q] - x[\alpha(q-p)+p]}{1-\alpha} = \nabla x[q]^T (q-p) = 0. \tag{28}$$

That is, the gradient in the patch should align with the vector originating at p.

2. **Gradients**

Show that the family of all images x that can be written as a function of only the direction of $q-p$

$$x[q] = f\left(\frac{q-p}{|q-p|}\right) \tag{29}$$

for any function f, satisfies the property below for a fixed pixel p

$$\nabla x[q]^T (q-p) = 0 \qquad \forall q. \tag{30}$$

Solution We apply the definition of the property and obtain

$$\nabla x[q]^T(q-p) = \nabla f\left(\frac{q-p}{|q-p|}\right)^T \frac{|q-p|^2 I_d - (q-p)(q-p)^T}{|q-p|^3}(q-p) = 0 \quad \forall f. \tag{31}$$