How the Brain Estimates Distances, Heights and Velocities, with Computer Vision Applications

Vincent Granville, Ph.D.
vincentg@MLTechniques.com
www.MLTechniques.com
Version 1.0, December 2024

A moving plane or car far away from you seems to move slowly and appears to accelerate as it gets closer to you, even though its speed is constant. Yet you are able to tell the approximate speed at all times, whether the vehicle is 100 yards or a mile away. Likewise, when looking at a skyscraper, at ground level, 100 feet away from the building, the windows in the upper floors seem smaller than those at the bottom, even though they are of the same size.

While I don't know exactly how the brain makes the corrections, in my case it seems that it uses a rather small mapping table between perceived and actual distances, using interpolation methods to increase accuracy. The purpose of this article is to explain the elegant mathematical formula used to generate that table. In the case of the human brain (mine at least), the table is built over time via reinforcement learning, and could be empty at birth: information retrieval to get the correct distance is just a memory call enhanced with simple interpolation, but not involving advanced mathematics.

1 Apparent versus real distance

Due to perspective, the perceived distance to an object is $1/d$, where d is the real distance. Let's now look at a the height H of a skyscraper, seen from a distance d at ground level. Based on Figure 1, we have $H = d\tan\theta$.
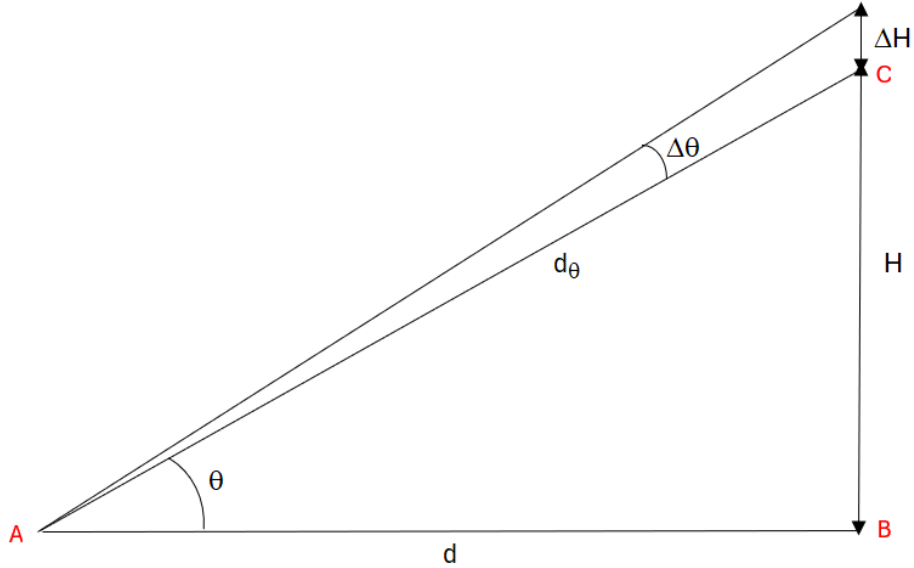

Figure 1: Height H of segment BC measured from A at ground level

The actual height H is a large sum of tiny increments ΔH, each proportional to the constant d. To measure H, think of it as being in an elevator located at A, moving at constant speed, and counting the milliseconds until the horizontal line of length d between you and the building, does not hit the building anymore when your elevation is greater than H. Now, let's turn to polar coordinates.

For the perceived height h_r, you are stuck at location A (the origin), looking up at various angles θ from bottom to top. Each angle θ contributes a tiny height $\Delta h_r(\theta)$ to the total perceived height. Using the weight $w(\theta) = \lambda/d_\theta$ where $\lambda > 0$ is a constant, and the notation $q(\theta) = d/\cos^2\theta$, we can write H and h_r as two versions of the same integral: the former with a fixed weight, and the latter with a variable weight, as follows:

$$H = \int_0^\theta q(\theta)d\theta = d\tan\theta, \quad h_r = \int_0^\theta w(\theta)q(\theta)d\theta.$$

The choice of $w(\theta)$ is dictated by the fact that the perceived distances are inversely proportional to the actual ones (the d_θ's) with a distinct d_θ for each θ. Also, $q(\theta) = 1/\cos^2\theta$ is the only continuous function that integrates to $H = d\tan\theta$ for all $0 \le \theta < \pi/2$; you cannot choose another one. Finally, since $d_\theta = d/\cos\theta$, we also have $w(\theta)q(\theta) = \lambda/\cos\theta$. Thus, the perceived height at angle θ is

$$h_r(\theta) = \lambda \int_0^\theta \frac{1}{\cos\theta} d\theta = \lambda \log\left[\tan\left(\frac{\pi}{4} + \frac{\theta}{2}\right)\right], \qquad 0 \le \theta < \frac{\pi}{2}, \tag{1}$$

where $\lambda > 0$ is a constant not depending on θ. The minimum value is $h_r(0) = 0$, while the maximum is infinite and reached when $\theta = \pi/2$. Since $H = d\tan\theta$, the perceived height $g_r(H)$ – defined as a function of the real height H – is equal to

$$g_r(H) = h_r(\theta) = h_r\left(\arctan\frac{H}{d}\right) = \lambda \log\left[\tan\left(\frac{\pi}{4} + \frac{1}{2}\arctan\frac{H}{d}\right)\right]. \tag{2}$$

With simple trigonometry, it is easy to simplify Formula (2) to obtain (3), where $\sinh^{-1}$ is the inverse hyperbolic sine function. From there, by inverting (3), one can express the actual height H as a function of the perceived height $g_r(H)$, see Formula (4). Here, g_r is defined as a function of H while h_r is a function of θ.

$$g_r(H) = \lambda \cdot \sinh^{-1}\left(\frac{H}{d}\right), \tag{3}$$

$$H = d \cdot \sinh\left(\frac{g_r(H)}{\lambda}\right). \tag{4}$$

Now is it easy to answer the following question. For a skyscraper seen from a distance d, what is the required height H' for the building to appear μ times taller, given that its real height is H? Let $\rho_\mu(H) = H'/H$ be the multiplier, also denoted as $\rho_\mu(H, d)$. It does not depend on λ. To obtain H', you need to solve the equation $g_r(H') = \mu g_r(H)$ where H' is the unknown. Using (3), the solution is

$$H' = d \cdot \sinh\left(\mu \cdot \sinh^{-1}\left(\frac{H}{d}\right)\right). \tag{5}$$

I computed some values of the multiplier $\rho_\mu(H)$ for various μ and H, assuming $d = 1$. See Table 1. Obviously, $\rho_\mu(H) = 1$ when $\mu = 1$, and $\rho_\mu(H) \approx \mu$ when H is small compared to d. When point A is close to B in Figure 1 and H is large, that is when θ is close to $\pi/2$, then $\rho_\mu(H)$ grows very fast as μ increases. All this is confirmed by personal visual experiences.

μ	H	$\rho_\mu(H)$	μ	H	$\rho_\mu(H)$	μ	H	$\rho_\mu(H)$	μ	H	$\rho_\mu(H)$
0.50	0.50	0.49	0.50	1.00	0.46	0.50	2.00	0.39	0.50	3.00	0.35
1.00	0.50	1.00	1.00	1.00	1.00	1.00	2.00	1.00	1.00	3.00	1.00
1.50	0.50	1.57	1.50	1.00	1.74	1.50	2.00	2.15	1.50	3.00	2.54
2.00	0.50	2.24	2.00	1.00	2.83	2.00	2.00	4.47	2.00	3.00	6.32
2.50	0.50	3.03	2.50	1.00	4.47	2.50	2.00	9.23	2.50	3.00	15.71
3.00	0.50	4.00	3.00	1.00	7.00	3.00	2.00	19.00	3.00	3.00	39.00

Table 1: Values of $\rho_\mu(H)$ depending on μ, H, with $d = 1$

Table 1 should be interpreted as follows. Assume that you are at ground level, 100 feet away from a 300 feet tall building ($H = 3$). The builders keep adding new floors and when you visit again, the building appears to be 1.5 taller ($\mu = 1.50$) seen from the same vantage point. In fact, this is an illusion due to perspective: in reality, it is actually 2.54 taller since $\rho_{1.50}(3) = 2.54$. The real height is now $762 = 2.54 \times 300$ feet. It is interesting to note that if H/d is an integer and μ an odd integer, then $\rho_\mu(H, d)$ is an integer. The easy proof relies on (5) and the fact that $\rho_\mu(H, d) = H'/H$, leading to the following result:

$$\rho_\mu(H, d) = \frac{d}{2H}\left[\left(\frac{\sqrt{H^2 + d^2} + H}{d}\right)^\mu - \left(\frac{\sqrt{H^2 + d^2} - H}{d}\right)^\mu\right]. \tag{6}$$

Formula (6) is a direct consequence of the definitions of the hyperbolic sine function and its inverse, combined with simple manipulations. From there, we obtain the linear recurrence relation

$$\rho_{\mu+1}(H, d) = \left(2 + \frac{4H^2}{d^2}\right)\rho_{\mu-1}(H, d) - \rho_{\mu-3}(H, d), \tag{7}$$

with initial conditions

$$\rho_0(H, d) = 0, \quad \rho_1(H, d) = 1, \quad \rho_2(H, d) = \frac{1}{d}\sqrt{4H^2 + 4d^2}, \quad \rho_3(H, d) = \frac{1}{d^2}(4H^2 + 3d^2).$$

2

2 Computer vision problems

The article *How can we Explain Perspective Calculations Simply?*, available here, provides an introduction. See also *The Size of an Object at a Distance* posted here by the Department of Homeland Security and intended for coast guards. The Wikipedia entry about angular diameter, here, shows applications in astronomy.

  If you rotate the triangle in Figure 1 by 90° to the right, then the discussion remains the same with height replaced by length, as in Figure 2. Despite appearances in the picture, railroad ties have the same width and are equally spaced. So, the distortions due to perspective and corrected by our brain are not illusions but also captured by any vision system. They can be accurately measured. In Figure 2, we do not have the context to determine the exact lengths, but the methodology presented here allows you to obtain correct ratios. Likewise, Formula (1) has a scale parameter λ that cannot be estimated based on the picture alone, yet relative lengths can be computed exactly.

Figure 2: Railroad ties are equally spaced despite appearances

  To measure velocities, get the location of a moving locomotive at two different times, compute the apparent distance $g_r(H)$ between the two locations on the picture, then get the real distance H with Formula (4). In this case, d is the distance between your vantage point above ground, and the railroad; λ is the scale parameter. You can use the same method to measure accelerations, and get it implemented in real-time.

  I conclude with the following observation. If you double the angle θ from 30° to 60° in Figure 1, then the actual height H is multiplied by 3, regardless of d. However the perceived height is multiplied by only 2.3975, regardless of λ. The exact value is $\log(2+\sqrt{3})/\log\sqrt{3}$, computed according to Formula (1). Here $\sqrt{3} = \tan(60°)$ and $2 + \sqrt{3} = \tan(75°)$.