The Physics Teacher

PAPERS | OCTOBER 01 2025

Inferring Trajectories of Floor-Bound Objects Using Video Analysis

David Kordahl

Phys. Teach. 63, 568–571 (2025)
https://doi.org/10.1119/5.0219972

Inferring Trajectories of Floor-Bound Objects Using Video Analysis

David Kordahl, Centenary College of Louisiana, Shreveport, LA

It is not possible in general to infer the full three-dimensional (3D) trajectory of an object from a single two-dimensional (2D) video, but it is possible in special cases. Video analysis is often used in physics education to infer trajectories of objects moving in planes that are roughly parallel to that of the camera's image plane. This article shows how it is possible to infer the trajectory of an object moving in a plane perpendicular to the image plane. The trajectory of an object moving around on the floor is reconstructed, correcting for projection, and an explicit algorithm for this process is given, with equations that allow direct calculation of floor coordinates from apparent video coordinates.

Introduction

For the past few decades, physics educators have routinely used video analysis to explore the trajectories of objects traveling in a single plane. The problem of projection errors for objects that stray outside that plane has been discussed under names such as "perspective correction"[1] and "parallax error,"[2] though the standard council usually boils down to the advice that one should back up as far as possible before taking a video.[3]

We can easily see why this is good advice via a negative example. Figure 1 shows a partial frame from a smartphone video of a toy car driving away from the camera with a nearly constant velocity (see video in the supplementary material). As the car recedes from the camera, its apparent velocity dramatically decreases. The complication, here, is that the plane of the car's motion is not parallel to the plane of the camera image, and light rays bouncing from the car into the camera have an ever-smaller angular displacement the further away the car travels.

Computer imaging experts have effectively solved the problem of how the 3D world can be projected into a 2D image,[4] but it is not possible in general to reconstruct the 3D world from a single 2D image.[5] However, there are exceptions. If we understand the particulars of our situation, including the position of our camera and its orientation relative to fixed surfaces, reconstruction may be possible. Figure 2 illustrates this idea schematically. For a pinhole camera, single points on the floor map to single points in the image plane. This suggests that if we have an unobstructed view of a surface whose geometry is known, the motion of objects moving along that surface may be reconstructed.

This article shows how to obtain the spatial trajectory of an object moving around on a flat floor from its apparent trajectory in a video. This analysis requires length references to be embedded in the video in a particular way and requires a few simple calculations from the video data. The next section derives these expressions in detail for the pinhole geometry, using ray tracing. Then, in the last section, this method is

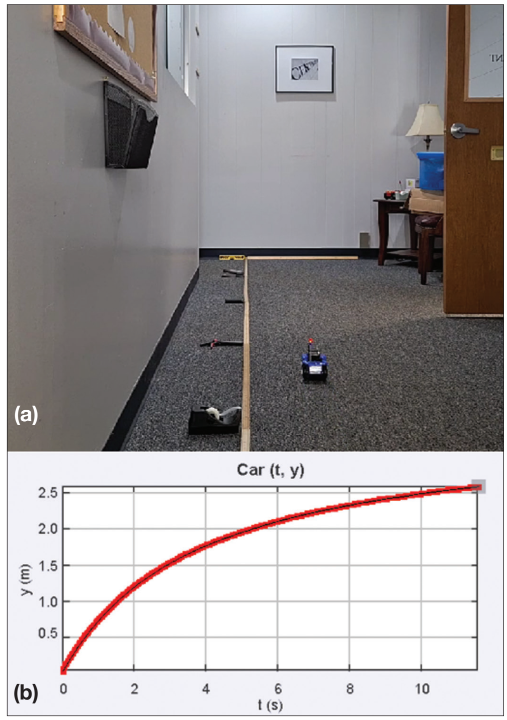

Fig. 1. In a video of a car traveling away from a camera at a nearly constant velocity (frame capture) (a), the *apparent* velocity of the car in the image will decrease as a function of time (plot from Tracker) (b). The length references coming toward the camera appear vertical in the camera image. The length reference on the far wall is parallel to the image plane.

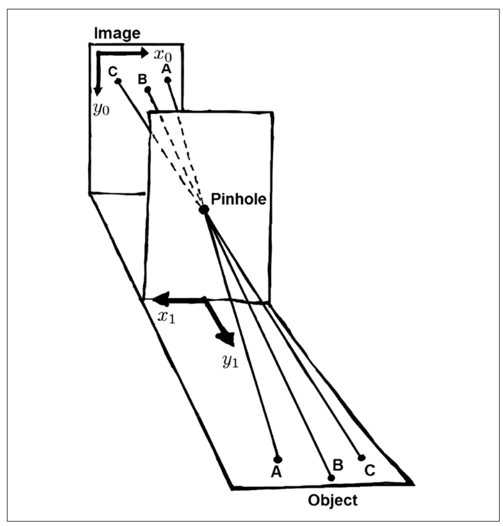

Fig. 2. There is a one-to-one relationship between points on the floor and points projected into the camera, as can be seen easily for the simple example of a pinhole camera. The "Projection correction" section works out how to correct projection errors for objects moving on the ground, guided by this physical model.

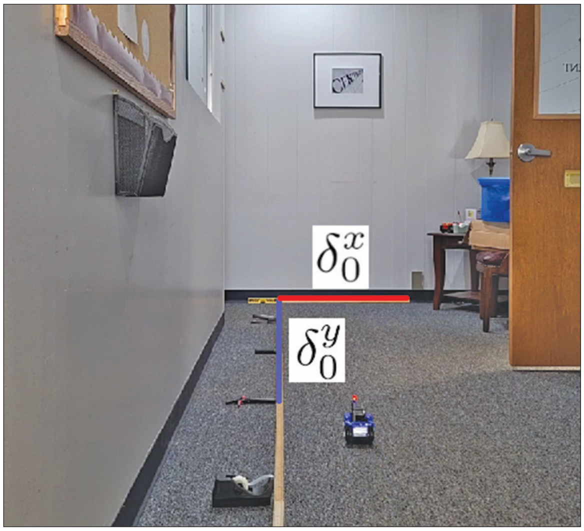

Fig. 3. The same still as in Fig. 1, but labeled with the reference lengths δ_0^x and δ_0^y. In this case, $\delta_0^x = \delta_1^x = 1.00$ m, while $\delta_1^y \neq \delta_0^y$, since $\delta_1^y = 3.00$ m while $\delta_0^y = 0.78$ m. The distance from the base of the camera to the horizontal reference δ_1^x is 6.0 m, so we can conclude that $\delta_1 = 6.0$ m. Since $\delta_0^x = \delta_1^x$, we can also conclude that $\delta_0 = 6.0$ m.

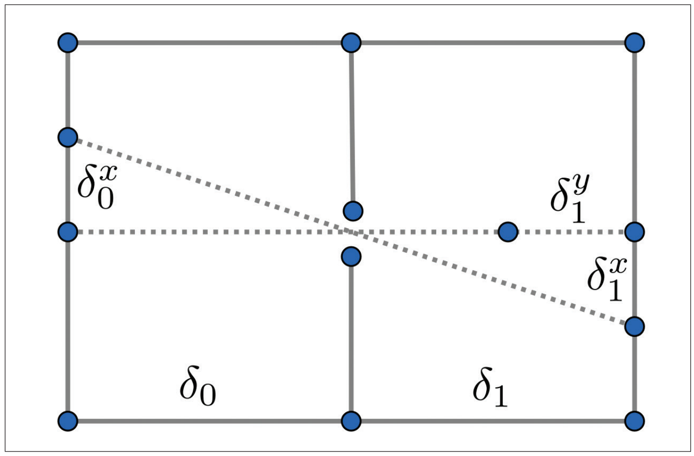

Fig. 4. Relating pinhole camera dimensions to physical dimensions—a "top-down" projection. The lengths δ_1 (from camera to horizontal length reference), δ_1^x (the horizontal length reference), and δ_1^y (the "vertical" length reference) are all measured physically. The "length" δ_0 is a fictitious value, but should equal δ_1 when $\delta_0^x = \delta_1^x$. In this figure, the x_0 coordinate increases going up, and the x_1 coordinate increases going down, since the pinhole camera inverts the image.

summarized and applied to the data shown in Fig. 1. Though this example is quite boring, it provides a useful first step for educators who are interested in expanding the possibilities of video analysis.

Projection correction

To simplify the derivation, we can press forward with the pinhole camera model of Fig. 2. Parameters can be extracted in terms of this model, allowing us to work backward from image to object coordinates.

Length references can be embedded in a sample video to make its geometry easier to interpret. Observe the meter sticks on the floor in Fig. 1. One stick has been deliberately positioned so that it appears exactly horizontal, and the five others have been positioned so that they appear to be exactly vertical in the image. This choice of floor coordinates has the camera with $x_1 = 0$ in the middle of the scene, aligned with the "vertical" length references. The $y_1 = 0$ coordinate designates the position directly below the camera. In this (x_1, y_1) Cartesian grid on the floor, x_1 values may be either positive or negative, but y_1 values are strictly positive, since negative y_1 values would be behind the camera.

There are six relevant lengths in this setup—three in the "object" space of the floor, and three in the "image" space of the camera. The "object" lengths are measured on the floor: the length δ_1 from below the camera to the horizontal length reference, the length δ_1^x of the horizontal length reference, and the length δ_1^y of the "vertical" length reference. Their counterpart lengths in the image are δ_0, δ_0^x, and δ_0^y. These lengths are illustrated in Fig. 3.

The horizontal length reference has the physical length δ_1^x, and the vertical length reference has the physical length δ_1^y. In Fig. 3, these lengths have been labeled as δ_0^x and δ_0^y, respectively, since the lengths in the picture are apparent lengths (subscript "0") rather than physical lengths (subscript "1"). The far wall, in this situation, is 6.0 m from the camera, so the length δ_1 is just 6.0 m, as is the distance δ_0 from the pinhole to image screen (though this value could be inferred, as discussed below). This hypothetical situation is three-dimensional, since light rays bouncing off the horizontal floor go through a pinhole and onto a vertical screen.

Figure 4 shows the "top-down" projection of the pinhole camera geometry, where items on the right represent items in the object space, and items on the left represent images in the pinhole camera. From this geometric setup, one can perform the ray tracing to relate object and image lengths. For instance, looking at Fig. 4, one can find δ_0 from similar triangles:

$$\frac{\delta_0^x}{\delta_0} = \frac{\delta_1^x}{\delta_1} \tag{1}$$

$$\rightarrow \delta_0 = \delta_1 \frac{\delta_0^x}{\delta_1^x}.$$

Though δ_0 is a fictitious length (since we are not, in fact, using a pinhole camera), it tells us about the image magnification (since, for a pinhole camera, magnification = δ_0/δ_1). Transparently, when $\delta_0^x = \delta_1^x$, then $\delta_0 = \delta_1$.

Next, we may look at the rays propagating in the vertical plane where $x_1 = 0$, containing the "vertical" length standard, as pictured in Fig. 5. In the camera coordinates, we may identify as (x_0^A, y_0^A) the point on the image of the vertical length standard that appears higher up (i.e., the place in the image picturing the point on the physical length δ_1^y that is farthest from the camera), and we may identify as (x_0^B, y_0^B) the point that appears lower down on the vertical length standard (i.e., the image of the point on δ_1^y that is nearest to the camera). We should find that

$$x_0^A \approx x_0^B, \text{ and } y_0^A > y_0^B.$$

A further analysis of Fig. 5 allows us to infer δ_z, the pinhole height, in terms of our length standards:

$$\delta_z = \frac{\delta_0^y}{\delta_1^y}\frac{\delta_1}{\delta_0}\left(\delta_1 - \delta_0^y\right). \tag{2}$$

This inferred pinhole height δ_z should approximate how high the physical camera aperture is from the ground.

Relating our camera coordinates (x_0, y_0) to the floor coordinates (x_1, y_1) can be simplified if we identify the midpoint y_0^C of our camera coordinates—where, as marked in Fig. 5, rays of light enter perpendicular to the pinhole. Suppose that we have obtained our observed length δ_0^y from the difference between y_0^A and y_0^B:

$$\delta_0^y = \delta_0^A - y_0^B. \tag{3}$$

One can then find the midpoint y_0^C in terms of y_0^A. Looking at similar triangles in Fig. 5, we can say that

$$\frac{y_0^C - y_0^A}{\delta_0} = \frac{\delta_z}{\delta_1}. \tag{4}$$

This can then be solved for y_0^C:

$$y_0^C = y_0^A + \frac{\delta_0\delta_z}{\delta_1}. \tag{5}$$

With all these parts in place, we can now calculate the mapping between the points in the video and the points on the ground. The sketches in Fig. 6 help to sharpen this intuition. Figure 6(a) is similar to Fig. 4, and allows us to determine, via similar triangles, that

$$\frac{x_0 - x_0^A}{\delta_0} = \frac{x_1}{y_1}. \tag{6}$$

Figure 6(b) is similar to Fig. 5, but it illustrates the vertical plane containing the point (x_1, y_1) as pictured in Fig 6(a). Again, from similar triangles,

$$\frac{y_0^C - y_0}{\sqrt{\left(x_0 - x_0^A\right)^2 + \delta_0^2}} = \frac{\delta_z}{\sqrt{x_1^2 + y_1^2}}. \tag{7}$$

Solving these expressions for x_1 and y_1, we can now reconstruct the physical coordinates as

$$x_1 = \frac{\left(x_0 - x_0^A\right)\delta_z}{\left(y_0^C - y_0\right)},$$
$$y_1 = \frac{\delta_0\delta_z}{\left(y_0^C - y_0\right)}. \tag{8}$$

These expressions can now be applied to video data. To simplify them, we may set $\delta_0 = \delta_1$ and $\delta_x^0 = \delta_x^1$; this has been done in Eq. (9), which makes calculations feasible using a spreadsheet.

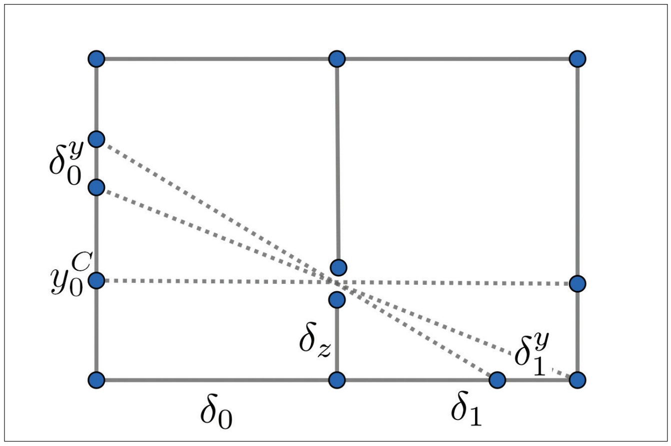

Fig. 5. Relating pinhole camera dimensions to physical dimensions—side view, in the plane where $x_1 = 0$. The lengths δ_1, from the camera to the wall, and δ_1^y, from the wall to the "vertical" scale marker, are measured physically. The parameter δ_0 has been determined by Eq. (1), and the inferred "pinhole height" δ_z (which approximates the physical height of the camera from the ground) is inferred from how foreshortened the observed length δ_0^y is in the image. To take account of image inversion, the y_0 coordinate increases vertically downward, and the y_1 coordinate increases to the right. The coordinate y_0^C marks where light rays entering parallel to the ground are imaged in this plane.

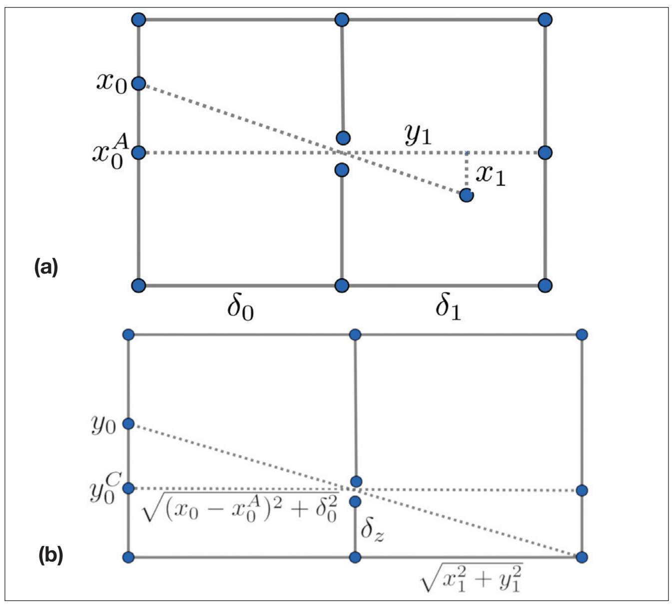

Fig. 6. Ray tracing for a general point (x_1, y_1) on the floor to a point (x_0, y_0) in the image. (a) A "top-down" projection, as in Fig. 4, showing where (x_1, y_1) on the ground will be mapped horizontally to a point x_0 in the image. (b) A vertical plane, as in Fig. 5. This plane contains (x_1, y_1), the pinhole, and the point $(x_1, y_1) = (0, 0)$ directly below the pinhole, which helps to determine the point y_0 in the image plane.

Correction example

This method can now be summarized, and everything can be put in terms of measured values. For an object traveling on the ground, we want to go from the apparent (x_0, y_0) coordinates in video data to infer the (x_1, y_1) coordinates for a Cartesian grid on the ground.

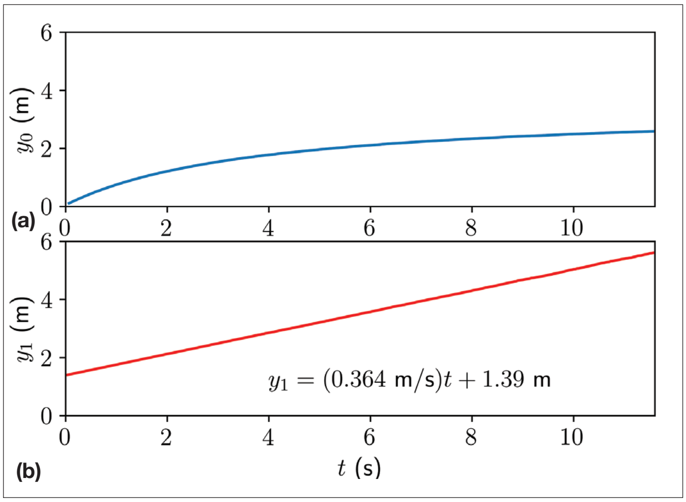

Fig. 7. Apparent position y_0 vs. t (a) of the toy car in the video featured in Figs. 1 and 3, and reconstructed position and y_1 vs. t (b).

To do this, we follow these steps:

1. Set up the camera without any tilt, and position perpendicular length standards, δ_1^x and δ_1^y, such that the two meet at a perpendicular corner at a distance δ_1 from the base of the camera. In the camera image, δ_1^x should appear horizontal, while δ_1^y should appear vertical but foreshortened, with apparent length δ_0^y (see Fig. 3).
2. Record a video of an object moving on the ground.
3. Extract apparent position (x_0, y_0) vs. time from the video using software (e.g., Tracker or Logger Pro), with the horizontal length standard δ_1^x setting the scale. Note the apparent video coordinates (x_0^A, y_0^A) of the end of the "vertical" length standard δ_y^1 that is farthest from the camera (i.e., the end whose distance from the camera base is δ_1).
4. Calculate the floor position (x_1, y_1) in terms of video positions (x_0, y_0):

$$x_1 = \frac{\delta_0^y\left(\delta_1 - \delta_1^y\right)\left(x_0 - x_0^A\right)}{\delta_0^y\left(\delta_1 - \delta_1^y\right) - \delta_1^y\left(y_0 - y_0^A\right)}, \tag{9}$$

$$y_1 = \frac{\delta_1\,\delta_0^y\left(\delta_1 - \delta_1^y\right)}{\delta_0^y\left(\delta_1 - \delta_1^y\right) - \delta_1^y\left(y_0 - y_0^A\right)}.$$

This analysis has been carried out for the position of the car in the video whose still was used in Figs. 1 and 3. The apparent vertical position of the car as a function of time that has been plotted in Fig. 7 is the same as that given directly by Tracker in Fig. 1, but once this data is analyzed using Eq. (9), it reveals a car whose recessional velocity is a constant 36 cm/s. This result matches both qualitatively and quantitatively with observations.

Though this analysis is simple, it immediately suggests possible applications. On the one hand, this analysis can be applied directly to objects moving across the floor—colliding hover pucks, say, or toy cars that can accelerate. On the other hand, the spirit of this analysis might also be applied more generally. If we can restrict the motion of an object to a 2D surface whose shape is well characterized, it should be possible to infer an object's trajectory in space from its apparent coordinates in a video. Generalizing this approach for a curved surface—say, for a coin rolling in a potential well—could present a worthwhile challenge for a motivated student. The article by Erdnüß[6] may help to bridge from the notation used here that of the computer vision experts.

Supplementary material

Readers can view the video in the supplementary material at *TPT Online*.

References

1. R. Allain, *Physics and Video Analysis* (Morgan & Claypool Publishers, 2016).
2. J. Stephens, M. Bostjancic, and T. Koskulitz, "A study on parallax error in video analysis," *Phys. Teach.* **57**, 193–195 (2109).
3. T. Martin, K. Frisch, and J. Zwart, "Systematic errors in video analysis," *Phys. Teach.* **58**, 195–197 (2020).
4. M. Pharr, W. Jakob, and G. Humphreys, *Physically Based Rendering: From Theory to Implementation*, 4th ed. (pbrbook.org, 2021).
5. R. Szeliski, *Computer Vision: Algorithms and Applications*, 2nd ed. (Springer, 2022).
6. B. Erdnüß, "Measuring in images with projective geometry," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XLII-1**, 141–148 (2018).

David Kordahl *teaches physics at Centenary College of Louisiana. He regularly writes about physics and other contentious topics for the website* 3 Quarks Daily.
dkordahl@centenary.edu

DOI: 10.1119/5.0219972