Questions

1. (a) When we discussed image registration, we mentioned a more general transformation (Shi and Tomasi) that is sometimes used for tracking:

$$I(\mathbf{x} + \mathbf{D}(\mathbf{x} - \mathbf{x_0}) + \mathbf{h}) = J(\mathbf{x}).$$

where $\mathbf{D}$ is a 2×2 deformation matrix. Write this transformation

$$\mathbf{x} \rightarrow \mathbf{x} + \mathbf{D}(\mathbf{x} - \mathbf{x_0}) + \mathbf{h}$$

as a homography.

   (b) Generalize this transformation so that it allows for any homography.

   (c) What are the advantages/disadvantages of using a more general model ?

2. Since a homography $\mathbf{H}$ maps points in homogeneous coordinates to points in homogeneous coordinates, you can multiply a homography by a constant without changing the mapping that it represents. This implies, in particular, that homographies have 8 degrees of freedom rather than 9. You might be tempted to represent these 8 degrees of freedom explicitly by insisting that one of the elements of $\mathbf{H}$ has a particular value, say $H_{33} = 1$. However, this is not always possible, for example, there are homographies for which $H_{33} = 0$.

   (a) $\mathbf{H}$ for which $H_{33} = 0$?
   Hint: think where the origin maps to.

   (b) We have seen an example of such a scenerio already in the course. What is it? (You need to answer (a) before you can do this one.)

3. (a) Consider a plane:

$$Z = Z_0 - Y \tan\theta$$

which we get by rotating the plane $Z = 0$ by θ degrees about the X axis and then translating by $(0, 0, Z_0)$. We parameterize points on the rotated plane by (s, t) such that the origin $(s, t) = (0, 0)$ is mapped to the 3D point $(x, y, z) = (0, 0, Z_0)$, and the s axis is in the direction of the X axis.

What is the homography $\mathbf{H}$ mapping $(s, t, 1)$ to the image projection plane $Z = f$, under perspective projection?

Hint: write the sequence of transformations that together map $(s, t, 1)$ to image projection plane coordinates (not pixel coordinates, since the question doesn't specify camera internal parameters).

   (b) For the plane in (a), what is the horizon ("vanishing line") in the image?
   Hint: the question is asking where points at infinity $(s, t, 0)$ map to.

4. Suppose the scene consists of a "ground plane" which in camera coordinates is

$$Y = h \ < 0.$$

On this ground plane, let's assume we have a regularly spaced grid of square tiles, i.e. tiles on a floor. These define two sets of parallel lines. These parallel lines need not be aligned with the camera's **X** and **Z** unit vectors, but rather can be rotated by some angle θ relative to these unit vectors.

Let the origin of the ground plane coordinate system be some 3D point (X_0, h, Z_0).

(a) What is the homography taking ground plane coordinates $(s, t, 1)$ to image plane coordinates, assuming a projection plane $Z = f$?

Hint: As in the previous question, you'll need to write out a sequence of transformations.

(b) What are the vanishing points that correspond to letting $s \to \infty$ (for finite t) and $t \to \infty$ (for finite s) ?

Answers

1. (a) The transformation

$$\mathbf{x} \to \mathbf{x} + \mathbf{D}(\mathbf{x} - \mathbf{x_0}) + \mathbf{h}$$

can be written using homogenous coordinates as

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 + D_{11} & D_{12} & h_x - D_{11}x_0 - D_{12}y_0 \\ D_{21} & 1 + D_{22} & h_y - D_{21}x_0 - D_{22}y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

where the 3×3 matrix is invertible, hence it is a homography.

(b) The transformation referred to in the previous question is not as general as it could be since the first and second elements of the third row are 0, and the third element is 1. To make the transformation model general enough to include all homographies, we could use

$$I(\mathbf{Hx}) = J(\mathbf{x})$$

where the $\mathbf{x}$ is written in homogeneous coordinates, and we could then try to find the $\mathbf{H}$ that minimizes the sum of squared errors over $\mathbf{x}$ in some local neighborhood of $\mathbf{x}_0$.

(c) The advantage of using a more general model is that it allows you to obtain a better fit to the data. The more restricted model fixes the value of some of the elements in the homography, in particular, the third row. By allowing these elements to vary when we find a least squares fit, we are guarenteed to get a better fit (or no worse, anyhow).

Another advantage is that by using a more general model, we sometimes allow ourselves to use a larger neighborhood. For example, a pure 2D translation model might be fine for a very small neighborhood, but if there is a stretching or shearing in addition to translation then pure 2D translation would not capture this. Similarly, if the true deformation involves perspective viewing of a plane, then you would likely need the full homography to capture this deformation.

What are the disadvantages in using a more general model? First, using more parameters causes the computation to be more expensive – this can be significant when we perform the computation at each pixel. Second, the more general model might be inappropriate in some situations, in which case the extra parameters might just "fit the noise". For example, if the deformation is caused by a non-planar object viewed by two different cameras (hence there is a 3D translation between the camera positions), then the transformation might not be well explained by a homography.

2. (a) If $H_{33} = 0$, then we would have

$$\mathbf{H} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} H_{13} \\ H_{23} \\ 0 \end{bmatrix}$$

that is, the origin $(x, y) = (0, 0)$ is mapped to a point at infinity. More generally, any 2D point (x, y) that is on the line $H_{31}x + H_{32}y = 0$ will be mapped to a point at infinity.

(b) An example is the following. Suppose we have two planes. The first plane is an image projection plane and the second plane is a scene plane, and $\mathbf{H}$ maps from the image plane coordinates to the scene plane coordinates. The mapping takes the origin of the image plane to a point at infinity on the scene plane. This means that the origin in the image plane lies on the vanishing line of the scene plane i.e. the origin lies on the horizon. An example is the scene ground plane $Y = -h$.

3. (a) The transformation from $(s, t, 1)$ to points on the rotated 3D plane (rotated about X axis), and then on to points on the image plane is:

$$\begin{bmatrix} wx \\ wy \\ w \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s \\ t \\ 1 \end{bmatrix}$$

so the homography is:

$$\mathbf{H} = \begin{bmatrix} f & 0 & 0 \\ 0 & f\cos\theta & 0 \\ 0 & \sin\theta & Z_0 \end{bmatrix}.$$

(b) The hint is to consider where the points at infinity $(s, t, 0)$ map to.

$$\mathbf{H} \begin{bmatrix} s \\ t \\ 0 \end{bmatrix} \equiv \begin{bmatrix} wx \\ wy \\ w \end{bmatrix}$$

In the image, these points (x, y) define the *horizon*.

$$\begin{bmatrix} s \\ t \\ 0 \end{bmatrix} \equiv \mathbf{H}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

The inverse is easily calculated (by hand) to be:

$$\mathbf{H}^{-1} = \begin{bmatrix} \frac{1}{f} & 0 & 0 \\ 0 & \frac{1}{f\cos\theta} & 0 \\ & & \\ 0 & -\frac{\tan\theta}{Z_0 f} & \frac{1}{Z_0} \end{bmatrix}$$

So which points (x, y) lie on the horizon? We want to know which points are mapped to points at infinity, that is,

$$(0, -\frac{\tan\theta}{Z_0 f}, \frac{1}{Z_0}) \cdot (x, y, 1) = 0$$

In particular,

$$y = \frac{f}{\tan\theta}.$$

Notice that if the surface is not slanted at all ($\theta = 0$), then $\tan\theta = 0$ and so the horizon is itself the set of points at infinity in the image plane i.e. $(x, y, 0)$.

4. (a)

$$\begin{aligned} \begin{bmatrix} wx \\ wy \\ w \end{bmatrix} &= \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & X_0 \\ 0 & 0 & h \\ 0 & 1 & Z_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s \\ t \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} f\cos\theta & 0 & fX_0 \\ 0 & 0 & fh \\ \sin\theta & \cos\theta & Z_0 \end{bmatrix} \begin{bmatrix} s \\ t \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} f(s\cos\theta - t\sin\theta + X_0) \\ fh \\ s\sin\theta + t\cos\theta + Z_0 \end{bmatrix} \end{aligned}$$

ADDED Nov. 12: Alternatively the first line could be written:

$$\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & X_0 \\ 0 & 1 & 0 & h \\ 0 & 0 & 1 & Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s \\ t \\ 1 \end{bmatrix}$$

(b) Consider lines that are parallel to either the s or t axes. First, let's get rid of the scale factor w. This gives:

$$(x, y) \; = \; (\frac{f(s\cos\theta + t\sin\theta + X_0)}{s\sin\theta - t\cos\theta + Z_0}, \frac{fh}{s\sin\theta - t\cos\theta + Z_0})$$

The two vanishing points defined as follows: Fix s and let $t \rightarrow \infty$:

$$\lim_{t\rightarrow\infty}(x, y) \;=\; (-f\frac{\sin\theta}{\cos\theta}, 0)$$

or, fix t and let $s \rightarrow \infty$:

$$\lim_{s\rightarrow\infty}(x, y) \;=\; (f\frac{\cos\theta}{\sin\theta}, 0).$$

Notice that the vanishing points lie *in the image* along the horizontal line $y = 0$.

Varying θ changes the positions of the vanishing points on the horizon. However, changing the origin position (X_0, h, Z_0) within the ground plane (keeping the y coordinate at h) does not change the position of the vanishing points.

