Computer Vision

CS 512

Homework 1

Russi Sinha

A20411286

Assignment 1

1)

a) $f = 10$ $\quad p = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{z} \begin{bmatrix} f & 0 \\ 0 & f \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$= \frac{1}{1} \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} 30 \\ 20 \end{bmatrix}$$ → coordinates of the projection

b) ~~Model~~ Camera model where the image plane is behind the COP ①, the image obtained is inverted as compared to the ~~image~~ camera model where the image plane is in front of the COP ②.
① is the better representation but ② is computationally good because it negates the inversion of the image.

c) The dimensions of the image are directly proportional to the focal length. So, if the focal length increases the projection of an object increases in size.

The dimensions of the image are inversely proportional to the distance to the object. So, if the distance to the object gets bigger the projection gets smaller.

d)

Point in 2D = (1, 1)

Point in 2DH = (1, 1, 1)

Another 2DH point that corresponds to the same 2D point = (2, 2, 2)

e) Point in 2DH = (1, 1, 2)

Corresponding 2D point = (0.5, 0.5)

f) The 2DH point (1,1,0) means that the corresponding 2D point is at infinity ~~becaus~~ because if we try to homogenize the 2DH point it would lead to infinity

Points with last coordinate as '0' represent direction

g) The non-linear projection equation is written as a linear equation in homogeneous coordinates by ~~mult~~ multiplying the projection matrix to a 3D point.

$$\underset{\text{2DH}}{\begin{bmatrix} U \\ V \\ W \end{bmatrix}} = \underset{\text{projection matrix}}{\begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix}} \underset{\text{3D}}{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}$$

$$U = fx, \quad V = fy, \quad W = z$$

which is the linear equation in homogeneous coordinates

h) $M = K[I|0]$

Dimensions of $M = 3\times4$

$$K = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad 0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Dimensions $\rightarrow$ (3×3) (3×3) (3×1)

j) $M = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 1 & 2 & 1 & 2 \end{bmatrix}$ $p = [1, 2, 3]$

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 1 & 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 18 \\ 48 \\ 10 \end{bmatrix}$$

Therefore the 2D point $p = (1.8, 4.8)$

2)

a) $(x, y) = (1, 1)$

Translating using transformation matrix

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}$$

Therefore $(3, 4)$ is the point after translation

b) $(x, y) = (1, 1)$

Scaling using transformation matrix

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$$

Therefore $(2, 2)$ is the point after scaling.

c) $(x, y) = (1, 1)$

Rotating by 45 degrees

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \cos 45 & -\sin 45 & 0 \\ \sin 45 & \cos 45 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 2/\sqrt{2} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{2} \\ 1 \end{bmatrix}$$

Therefore $(0, \sqrt{2})$ is the point after rotating by $45°$.

d) $(x, y) = (1, 1)$

Rotating by 45° at $P(2,2)$

Step1: Translating P to origin

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

Step 2: Rotate by 45°

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos 45 & -\sin 45 & 0 \\ \sin 45 & \cos 45 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ -2/\sqrt{2} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\sqrt{2} \\ 1 \end{bmatrix}$$

Step 3: Translating origin back to P

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -\sqrt{2} \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 2-\sqrt{2} \\ 1 \end{bmatrix}$$

Therefore $(2, 0.586)$ is the point after rotating.

e) Let P be the matrix of the object to be transformed

Rotating the object using matrix R

$$= RP$$

Translating the result using matrix T

$$P' = TRP$$

where P' is the resulting matrix after both the transformations.

(f) Given $M = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

We know that the 2D transformation matrix in homogeneous coordinate for scaling is

$$\begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ which is similar to M

where $S_x = 3$ and $S_y = 2$

Therefore, the matrix M scales the point P by (3, 2)

g) Given $M = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$

We know that the 2D transformation matrix in homogeneous coordinate for translation is $\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$ which is similar to M wher $t_x = 1$ and $t_y = 2$

Therefore, the matrix M translates the point P by (1, 2)

h) $M = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

We ~~Kno~~ know that M is the ~~transfor~~ transformation matrix for scaling (Question 2.f.). The matrix that will reverse the effect of M is the inverse of M

$$M^{-1} = \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

i)

$$M = R(45)\,T(1,2)$$

$$M^{-1} = \left(R(45)\,T(1,2)\right)^{-1}$$

$$= T^{-1}(1,2)\,R^{-1}(45)$$

$$= T(-1,-2)\,R(-45)$$

j) Two vectors are perpendicular to each other if their dot product is zero

Given $A = (1, 3)$

Let $B = (x, y)$ be the vector perpendicular to A

$A \cdot B = 0$

$(1, 3) \cdot (x, y) = 0$

$x + 3y = 0$

$x = -3y$

Let $y = 1$, then $x = -3$

Therefore $(1, -3)$ is perpendicular to $(1, 3)$

k) Given $a = (1, 3)$, $b = (2, 5)$

Projection of a onto direction of b

$$\mathrm{Proj}_b(a) = \frac{a \cdot b}{|b|^2} b = \frac{17}{29}\begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$= \left(34/29, 85/29\right)$ is the projection of $(1, 3)$ onto direction defined by $(2, 5)$

3)

a) The general projection matrix is a transformation matrix ~~use~~ which is used to move from 3D to 2D coordinate system. Different coordinate systems are used to conduct the transformation because the ~~the~~ real world object must be translated to the camera and then translated into an image

b) When the camera is rotated by R and translated by T w.r.t. world

$$M_{c\leftarrow w} = \tilde{R}^{-1}\tilde{T}^{-1}$$

$$= \left[\begin{array}{c|c} R & 0 \\ \hline 0 & 1 \end{array}\right]^{-1} \left[\begin{array}{c|c} I & T \\ \hline 0 & 1 \end{array}\right]^{-1}$$

$$= \left[\begin{array}{c|c} R^T & 0 \\ \hline 0 & 1 \end{array}\right] \left[\begin{array}{c|c} I & -T \\ \hline 0 & 1 \end{array}\right]$$

$$= \begin{bmatrix} R^T & -R^T T \\ 0 & 1 \end{bmatrix}$$

$$= \left[\begin{array}{c|c} R^* & T^* \\ \hline 0 & 1 \end{array}\right]$$ where R^* and T^*

are rotation and translation of world w.r.t. camera.

c) The rotation matrix describing the rotation of the camera w.r.t. world is

$$\begin{bmatrix} - \hat{x}_c^T - \\ - \hat{y}_c^T - \\ - \hat{z}_c^T - \end{bmatrix}$$

where $\hat{x}_c$, $\hat{y}_c$ and $\hat{z}_c$ are the the unit vectors for the camera axes and $\hat{x}_c^T$, $\hat{y}_c^T$, $\hat{z}_c^T$ are the corresponding transpose of the vectors.

d) $$M = \begin{bmatrix} R^* & T^* \\ 0 & 1 \end{bmatrix}$$

R^* and T^* are the rotation and translation of the world w.r.t camera

e)

e) Given k_u, k_v pixels per mm in the x and y direction and optical center of camera is translated by $(u_0, v_0) = (512, 512)$ pixels, the transformation matrix that will convert camera coordinate to image coordinate is

$$M_{i \leftarrow c} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Substituting values for u_0 and v_0 we get

$$M_{i \leftarrow c} = \begin{bmatrix} k_u & 0 & 512 \\ 0 & k_v & 512 \\ 0 & 0 & 1 \end{bmatrix}$$

f) $M = K^* [R^* | T^*]$

K^* contains the ~~intrsic~~ intrinsic parameters i.e. focal length (f), scale in pixels/mm (k_u, k_v) and ~~transla~~ translation of optical centre of camera (u_0, v_0) (pixels) and $tg\,\alpha$, if skew exists.

R^* and T^* contain the extrinsic parameters which is the rotation and translation of the world w.r.t camera

g) When an image is scanned from the sensor, there is a small shift.
The skew is very tiny ~~which~~ but for the image to be 100% accurate the skew parameter needs to be added in the camera model.

h) If we take into account the radial lens distortion in the camera model, the image shrinks depending on the distance from the centre. It shrinks more as we move further away from the centre. This occurs in wide angle lenses.

Radial lens distortion can be included as follows:

$$p^{(i)} = \begin{bmatrix} 1/\lambda & 0 & 0 \\ 0 & 1/\lambda & 0 \\ 0 & 0 & 1 \end{bmatrix} K^{*} \left[R^{*} \mid T^{*} \right] \underline{P}^{(\omega)}$$

where $\lambda = 1 + k_1 d + k_2 d^2$

$d \rightarrow$ distance from centre

This introduces a complication since λ depends on the distance from the centre which is non-linear and can result in large values of λ at the edges of the image.

i) The weak perspective camera mode correts the vanishing point that appears in a regular camera ~~model~~ by making the imaging rays parallel

An affine camera is a model for a camera that is not real. It is used as a computational model since it is mathematically ~~basi~~ ~~a~~ ~~a~~ easier to ~~e~~ compute.

4)

a) Surface radiance is the light that is reflected from ~~a~~ a surface.
Image irradiance is the light that is captured at the sensor.

b) $E(p) = L(p) \frac{\pi}{4} \left(\frac{d}{f}\right)^2 (\cos\alpha)^4$

where $E(p)$ - image irradiance
$L(p)$ = surface radiance
d = diameter of lens
f - focal length
α = angle between the optical axis and surface normal

c) The albedo of a surface is the amount of light that is reflected from a surface

d) RGB color model is used to represent colors because the human vision works in ~~[illegible]~~ a similar way. They have sensors that are sensitive to red, blue, green and these can be combined in various proportions to get a wide ~~[illegible]~~ range of colours.

e) In the RGB cube, the colors along the line that connect (0,0,0) and (1,1,1) represent a gray scale.

f) The C.I.E. RGB model is used to map RGB colors to real world. It uses a graph that plots R, G, B values individually to a specific wavelength.

g) The luminance component Y is used to convert ~~...~~ any color image into grayscale since Y covers the widest wavelength
~~And this rep~~

h) The advantage of the LAB color space is that since it is ~~a~~ essentially a ~~...~~ non-linear mapping of the ~~...~~ XYZ color space, it makes up for the lack of human perception in the XYZ color space using the euclidean distance which is more representative of the perception distance.