cs512 Assignment 1: Review Questions

Amit Nikam
Department of Computer Science
Illinois Institute of Technology

October 8,2020.

---

1. Geometric image formation:

   a) Coordinates of the point p when projecting it onto the image are (30,20,1).
   b) The pinhole camera model where the image plane is behind the image center of projection corresponds better to a physical pinhole camera model. The alternative model where the image in front of the center of projection is used to better represent the model. In alternative model, no negative coordinates exist.
   c) As the focal length gets bigger, the image will get bigger, but might lose its quality after creation length. On the other hand as the distance gets bigger from the object, the object starts appearing smaller on image due to increased Z-axis.
   d) The coordinates of 2D point (1,1) in 2DH are (1,1,1). Other 2DH points that would correspond to the same 2D point are (2,2,2) (3,3,3) so on.
   e) For 2DH point (1,1,2) the corresponding 2D point would be (1/2 , 1/2).
   f) Such 2DH points represent direction and are called points at infinity.
   g) The projection equation happens to be non-linear due to the z-axis division in 2D form. But when we convert to homogeneous coordinates, the z-axis becomes linear along with x and y axis.
   h) M is 3x4 projection matric. K is 3x3 internal parameters. I is 3x3 identity matrix. 0 is 3x1 zero matrix.
   i) The coordinates of the 2D point p obtained would be (1.8 , 4.6).

2. Modeling Transformations:

   a) The coordinates after translation using transformation matrix will be (3,4).
   b) The coordinates after scaling using transformation matrix would be (2,2).
   c) The coordinates after rotating point (1,1) by 45 degrees are (0,√2).
   d) The coordinates after rotating point (1,1) by 45 degrees about the point (2,2) are (2,2-√2).
   e) Combined matrix to be applied would be TR.
   f) The point p will get scaled 3x in x-axis and 2x in y-axis.
   g) The point p will get translated by 1 in x-axis and 2 in y-axis.
   h) The transformation matrix that will reverse the effect is

      [1/3, 0, 0],

      [0, 1/2, 0],

      [0, 0, 1].

i) Inverse of the transformation matrix will be M=T(-1,-2)R^T(45).

j) Vector perpendicular to the vector (1,3) is (3,-1).

k) The projection is (34/29 , 85/29).

3. General camera model:

a) A general projection matrix allowed us to represent a 3D world object in more than just a camera centered coordinate system i.e image. It enables for transformation between world and camera coordinate systems by aligning the camera w.r.t world.

b) Transformation matrix $M_{C<W}$=

$$\begin{bmatrix} R^* & | & T^* \\ 0 & | & 1 \end{bmatrix}$$

c) Rotation matrix describing the rotation of camera w.r.t world is $[x^T,y^T,z^T]^T$

d) R* is R^T which is inverted rotation on world coordinates. T* is -R^TT which is product of inverted translation and rotation on world coordinates i.e '-T' and 'R^T'.

e) $M_{I<C}$= $\begin{bmatrix} k_u, & 0, & 512 \\ 0, & k_v, & 512 \\ 0, & 0, & 1 \end{bmatrix}$

f) K* contains the intrinsic parameters such as focal length, scaling and translation. [R*|T*] contains the extrinsic parameters with rotation and translation of world w.r.t camera coordinates.

g) The images are a little skewed in camera model, the skewness is very very small but becomes important to be fixed for certain applications where precision is needed.

h) Radial lens distortion is when a distortion is introduced to the camera model because of the nature of the lens, specially wide view lenses that makes straight lines appear curved in the image. This introduced the challenge of finding the linear distortion coefficient and quadratic distortion coefficient to fix it w.r.t the distance of point from center.

i) Weak perspective camera is the one where we don't see perspective i.e. there is no foreshortening. AN affine camera is a computational model that is a 3x4 matrix with 2x4 being arbitrary numbers and the last row being [0,0,0,1].

4. Color and photometric image formation:

a) Surface radiance is light in the scene measured as power of light per unit area reflected from surface. Image irradiance is light in the image measured as power of light per unit area reflected at the image.

b) $E(p) = L(p)\ (pi/4)\ (d/f)^2\ (\cos\alpha)^4$; where E(p) is light at image, L(p) is light at surface, d is diameter of lens, f is focal length and cos α is angle between principal axis and surface normal.

c) Albedo of a surface is its reflection coefficient. It can be between [0,1] both inclusive.It is closer to 1 for highly reflective surfaces and 0 for surface that reflect no light.

d) Human vision is made up of RGB receptors at retina. Since human brain makes the different colors using RGB, it makes sense to use RGB to represent colors since it will come naturally to human vision.

e) The colors along the line that connects (0,0,0) and (1,1,1) are different variants of white and black colors mix, so it can be seen as different shades of gray(from darkest to lightest).

f) Real-world colors are mapped to RGB through empirical tables obtained through manual experiments done by CIE Labs. These RGB spectrum had negative values so they were converted to different model with positive values.

g) Luminance component Y is the intensity. It is useful when working one grayscale images, as only taking Y is required to process the images.

h) Euclidean distance in RGB space does not correspond to human perception but in LAB color space it does correspond to human perception.

5. Noise and filtering:

a) SNR is estimated by taking the ration of signal energy i.e. variance of signal and noise energy i.e variance of noise. SNR is smaller for noisy images. Signal energy is (1/number of pixels) x $\Sigma(I_{(i,j)}-I_{(mean)})^2$. Noise energy can be calculated by taking sample from image or using multiple images.

b) Gaussian noise is a statistical noise that has a probability distribution function of standard deviation. Impulsive noise also known as salt and pepper noise are random peaks of noise in image pixels. A median filter handles impulsive noise better as extreme ends are eliminated, where as in averaging filter the extremes can change the mean significantly.

c) Each pixel will have a value of 18 which is high intensity value. Smoothing this by diving 9 will give a value of 2.

d) The operation of getting derivative of a filter can be done more efficiently by using Gaussian filters with mean 0, this gives the same zero crossing outcome.

e) Three ways to handle boundaries during convolution are (1) zero-padding where zeros are considered at the boundaries (2) mirror/replication - in which the adjacent pixel is mirrored/replicated (3) ignore - in which the boundaries are ignored and so the image convoluted is smaller in size than the original image.

f) Sum of all the entries in this filter is 9, but smoothing it gives outcome of 1. The sum can selected as it is such that entries in the filter have higher impact on the outcome.

g) 1D convolution filters need to be applied twice on the image. After applying 1D convolution, apply the another 1D convolution to the first result. It is more efficient to implement two 1D convolutions. Yes, it is possible to implement 2D filters in this manner.

h) The size of this 1D Gaussian filter should be 11.

i) Gaussian image pyramid is useful for tackling aliasing images. It is also useful for multiple scale analysis of the images. Gaussian pyramid is produced by performing a Gaussian convolution on the image and sampling it. This image is input at next level and the process continues until last level.

j) Laplacian pyramid is useful for image compression. In Laplacian pyramid, first Gaussian convolution is performed on the original image Img_j at level J after which it is sampled, this gives the $Img_{(J-1)}$ as outcome which is input for next level. The $Img_{(J-1)}$ is then reverse sampled to original size and a Gaussian is performed on this image. Finally, the difference between this image and original image is taken as residuals which gives us loss between two consecutive levels. This process goes until last level.

6. Edge detection:

a) Edge detection is useful for detecting features in the image. It helps in localizing images. It is required to correspond to scene elements, be invariant(to illumination, scale, pose, viewpoint) and be a reliable detection.

b) Smoothing is needed to reduce the noise in the image without affecting the edges(noise can give bad edges). Edges need to be enhanced in order to make it easy to localize them.

c) Central difference filter and soble filter can be used to compute image gradient. Image gradient is the first derivative of an image which gives the degree of change between a pixel and its next pixel. It can be used to detect edges by finding derivatives that are more than a certain threshold.

d) Sobel filter takes a smoothing filter of size 2x2 and forward difference of size 2x2 w.r.t rows and takes their convolution. The result is a sobel filter for columns(vertical). Same is done with forward difference of columns to get sobel filter for rows(horizontal).

e) To generate a more accurate derivative filter with an arbitrary sigma, we convert the discrete filter to continues, takes its derivative and then convert it to discrete again for each of the dimensions of the image. 2^{nd} degree filter for sigma = 2 is [0.055 , 0.135 , 0.244 , 0.304 , 0.221 , 0 , -0.221 , -0.304 , -0.244 , -0.135 , -0.055].

f) In a first order derivative a threshold is used to localize the edges, where as in second order derivative zero crossing is used to detect edges and localize them.

g) For using LOG, it is applied to the images, then if the threshold of the resultant is more than a threshold, 1 is stored in its equivalent pixel position else a 0 is stored. Outcome is a binary matrix of image. Then finally we mark the edges at zero crossings. LOG is [0.0034 , 0 , $-\infty$, 0 , 0.0034].

h) In a standard edge detection, the image is processed first and final step is to identify edges. In Canny edge detection, the edge is identified during run-time using directional derivatives and then it is localized. The condition for detecting an edge candidate in Canny is that their gradient magnitude needs to be larger than the threshold value.

i) In Canny algorithm, we start scanning for edges using directional derivatives with a higher threshold value $\mathfrak{T}^H$. Once the edge is detected we use the lower threshold $\mathfrak{T}^L$ to scan edges in its local neighborhood. While doing non-maximum suppression of Canny algorithm, we take the local gradient of Smoothed Image $\nabla(I*G)$. Then we find the local maximum value in it along θ, where θ is angle(0,45,90,135, ...) and set that to 1, otherwise 0. After this we record the results in respective position in a different matrix that’s the same size of the image. This continues until the entire image i.e. each pixel value is set.