CS 512 HW4

1)

a) In computer vision applications, outliers typically occur within a sample (image) due to pixels that are corrupted by noise, alignment errors, or occlusion.

b) objective function for Robust estimation –

$$E(\theta) = \sum_{i=1}^{n} \left(d(x_i,\theta)\right)^2$$

or, $$E(\theta) = \sum_{i=1}^{n} S_\sigma\left(d(x_i,\theta)\right)$$

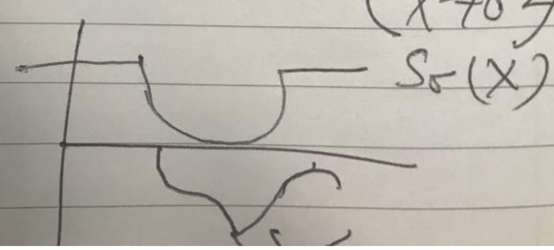
In standard least square error fitting, $S_\sigma(x) = x^2$

In Robust estimation,

$$S_\sigma(x) = \left(\frac{x^2}{x^2+\sigma^2}\right)$$

c) German-McClure function,

$$S_\sigma(x) = \left(\frac{x^2}{x^2+\sigma^2}\right)$$



d)

Random sample consensus (RANSAC) is an iterative method to estimate parameters of a mathematical model from a set of observed data that contains outliers, when outliers are to be accorded no influence on the values of the estimates.

The RANSAC algorithm is essentially composed of two steps that are iteratively repeated:

1. In the first step, a sample subset containing minimal data items is randomly selected from the input dataset. A fitting model and the corresponding model parameters are computed using only the elements of this sample subset. The cardinality of the sample subset is the smallest sufficient to determine the model parameters.
2. In the second step, the algorithm checks which elements of the entire dataset are consistent with the model instantiated by the estimated model parameters obtained from the first step. A data element will be considered as an outlier if it does not fit the fitting model instantiated by the set of estimated model parameters within some error threshold that defines the maximum deviation attributable to the effect of noise.

The number of points at each attempt should be small as small set will all inliers will give good points whereas large number of points will result in noise.

e)

Parameters of RANSAC algorithm can be given as,

n=Number of points to draw at each evaluation
d=Max. # points needed
w = number of inliers in data / number of points in data
t=The threshold value to determine when a data point fits a model

Formula to estimate number of trials can be given as,

$$k = \frac{\log(1-p)}{\log(1-w^n)}$$

where, w=probability that a point is inlier
p=probability that at least one exponent doesn't have outliers.

f)

In computer vision, image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

Agglomerative approach is for spatial clustering. It can be written as,

i) Start with each pixel in separate cluster.
ii) Merge clusters if distance between feature vectors is low.

Divisive(split) approach can be written as,

i) Start all in one cluster
ii) If distance between feature vectors is big then split clusters relatively.

g)

h)

Mean Shift Algorithm:

i) For each datapoint $x \in X$, find the neighbouring points *N(x)* of *x*

ii) For each datapoint $x \in X$, calculate the ***mean shift*** *m(x)* from this equation:

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x)x_i}{\sum_{x_i \in N(x)} K(x_i - x)}$$

iii) For each datapoint $x \in X$, update $x \leftarrow m(x)$.

iv) Repeat 1. for *n_iteations* or until the points are almost not moving or not moving

2)

a) Projection equation, p=MP

Forward Projection:
In forward projection we convert world coordinates to camera coordinates then camera coordinates are used to find image coordinates.

Calibration:
In calibration we project 3DH world coordinates to get 2D image points based on which we calculate camera intrinsic(k*) and extrinsic parameters(R* and T*).

Reconstruction:
In reconstruction we find our world points(3D) from given image points(2D).

From the above the easiest one is Forward Projection whereas difficult one is Reconstruction.

b)
Two necessary inputs for camera calibration are,

i) 2D Image points (xi,yi)

ii) 3D World Points (X,Y,Z)

(C) Steps in non-planar algorithm

(i) Estimate projection matrix M.

(ii) Find parameters (K^*, R^*, T^*)

In step one, we need to find the projection matrix M. With given 2D Image points and 3D world point.

Here $\underset{(2DH)}{p} = M \underset{(3DH)}{P}$

$M \rightarrow 3\times4$ Matrix
$\rightarrow K^*[R^*|T^*]$

d) Projection matrix. M

$$= \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 0 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$P_i = [1, 2, 3]$

P_i in Homogeneous,

$P_i = [1, 2, 3, 1]$

$$x_i = \frac{\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} [1\ 2\ 3\ 1]}{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [1\ 2\ 3\ 1]}$$

$$= \frac{1+4+9+4}{1+2+3+1}$$

$$= 18/7$$

and, $y_i = 14/7 = 2$

So, $(x_i, y_i) = (18/7, 2)$

e) World-image points:

$$(1,2,3) \rightarrow (100,200)$$

First two rows of projection matrix,

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 & 0 & 0 & -100 & -200 & -30 & -10 \\ 0 & 0 & 0 & 0 & 1 & 2 & 3 & 1 & -200 & -400 & - & -200 \end{bmatrix}$$

f) Minimal 6 points are necessary to be able to find a unique solution for m.

In order to find matrix m, we consider,

$$AX = 0$$

where, $A \rightarrow 2n \times 12$ matrix

$X \rightarrow 12 \times 1$ matrix

$0 \rightarrow 2n \times 1$ matrix

$$\begin{bmatrix} P_1^T & 0 & -x_1 P_1^T \\ 0^T & P^T & -y_1 P^T \\ \vdots & & \\ P_n^T & 0^T & -x_n P_n^T \\ 0^T & P_n^T & -y_n P_n^T \end{bmatrix} \underbrace{\begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}}_{(12 \times 1)} =$$

After taking value of A, we take SVD of A and take the last column of V to get the matrix.

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$(2n \times 1)$

8) To find the unknown parameters projection matrix M,

(i) Estimate M

(ii) Break M into R^*, K^* and T^*

$M \rightarrow S\hat{M}$ $\left[\because \hat{M} \rightarrow \text{estimated}, \; S \rightarrow \text{unknown} \right]$

So, from the above equation we can find S ~~by~~ as we have estimated M and exact M value.

h) Quality of projetion matrix can be given as,

$$\{P_i\}^n \rightarrow \{P_i\}$$

Image points → world points

For camera calibration, we will calcullate the estimate of parameters K^*, R^* and T^* where

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

$$E(K^*, R^*, T^*) = \sum_{i=1}^{n} \left(x_i - \frac{m_1^T P_i}{m_3^T P_i}\right)^2 + \left(y_i - \frac{m_2^T P_i}{m_3^T P_i}\right)^2$$

i) In order to perform planar camera calibration, we need to show the view to the camera for more than one time, minimum is three times to calibrate.

Approach ⇒

i) Estimate 2D homography between ~~a~~ calibration target and image.

ii) Estimate intrinsic paramet from several views.

iii) Compute extrinsic params.

In non-planar Calibration, it uses 3D image as Calibration target. So, we need to find the pixel coordinates of all the corners and based on these values camera calibration is done while in planar, a single plane picture is used with different views.