cs512 Assignment 2: Review Questions

Amit Nikam
Department of Computer Science
Illinois Institute of Technology

October 20,2020.

---

1. Corner detection:

   a) The basic principle in corner detection is (1) to find the correlation matrix of gradients in local window (2) Then find eigenvalues at correlation matrix (3) then detect corners in window if eigenvalues are sufficiently large. The number of principal directions is assessed based on number of directions in orientation histogram.

   b) To find principal directions using PCA we try to find a vector such that projections of gradient orientations in a local patch onto this vector is minimized.

   c) Correlation matrix that can be used for corner detection is =

   $$\begin{bmatrix} 4 & 6 \\ 6 & 44 \end{bmatrix}$$

   d) The condition is to consider the smallest eigenvalue of the gradient correlation matrix and to take the corresponding eigenvector as a solution.

   e) While performing non-maximum suppression for corner detection we (1) compute $\lambda1.\lambda2$ for all windows. (2) Select $\lambda1.\lambda2 > \Im$ (threshold) windows and sort in descending order. (3) Select the top of the list as corner and delete all other corners in its neighborhood. (4) We stop this process once detecting certain % of points as corners is done.

   f) Harris corner detection computes the determinant and trace of the correlation matrix for a window instead of computing the eigenvalues directly.

   g) The formula for computing better localization of a corner is $P^*=C^{-1}(\sum\nabla I_{(xi)}\nabla I_{(xi)}^T Xi)$. In this formula the objective is to get minimized location of corner P*. We get this by connecting each point Xi to P (lets say Xi-P) and projecting the gradient at Xi onto (Xi-P). The best P or P* is the one that minimizes the sum of all projections i.e. its a corner. The condition for the solution in the formula to exist is that there needs to be a corner in the window to be localized.

   h) While using HOG to characterize feature points, we first split each patch into cells which are possible overlapping. Then we create a orientation histogram in each cell, using edge or gradient directions, possible weighted by their distance form center of gradient magnitude. Then finally we concatenate the orientation histograms.
   A good characterized feature point must be translation invariant, rotation invariant, scale invariant and illumination invariant.

   i) In SIFT, the image gradients are divided into cells. For each cell, a keypoint descriptor(orientation histogram) is computed using weighted sum of gradients. Then the histogram is aligned based on dominant direction.

2. Line detection:

a) Slope can go to an ∞ so it has a problem with bounds. Also another problem is representing vertical lines.

b) A = $1/\sqrt{2}$, B = $1/\sqrt{2}$ and C = -10. We consider point $(15/\sqrt{2}, 5/\sqrt{2})$ on the line and find angle it makes with the origin as 18.5 degrees. So the new equation becomes $(15.\cos(18.5)/\sqrt{2})+(5.\sin(18.5)/\sqrt{2})$ = d. We get the new d as 11.18. To prove it holds true, we take the hypotenuse as $\sqrt{(15^2/2+5^2/2)}$ and get the same distance i.e 11.18, so the equation holds true.

c) The vote of each point in the image looks like a line in the parameter plane while using polar representation.

d) Each point in the image, we use the point coordinates to scan for A_{m-1} parameters of the parameter plane and case a vote for A_m. So for a line each point coordinate x,y scans for a θ and votes for a d which is represented as a line. Line intersections in parameter plane are parameters common for two different points in image space. The parameters having highest votes is selected as a model outcome.

e) If the bin size is larger in parameter space, more efficiency is achieved. While small bins give a more accurate outcome, but are not as efficient.

f) If the normal at each voting point is known, voting can be improved by taking the dot product of the normal and point coordinates to cast a vote.

g) There are three dimensions in parametric space while using Hough transform for circles.

3. Model fitting 1:

a) Disadvantage of using the equation y =ax+b for line fitting is that it only takes vertical errors i.e. algebraic distance. This is not optimal as there can be cases where the desired outcome of the model could be significantly different than the expected one. A vertical line cannot be fitted accurately using this model.

b) $L = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$

c) To fit a line using the explicit line equation $L^TP=0$, we minimize it such that we obtain the optimal unknown parameters L*. The equation that has to be solved for the unknown line parameters is $\nabla E(L)=0$ i.e. SL=0 where S is the correlation matrix and L is eigenvector belonging to zero or smallest eigenvalues of the correlation matrix S.

d) $S = \begin{bmatrix} 5 & 15 & 3 \\ 15 & 46 & 10 \\ 3 & 10 & 3 \end{bmatrix}$

e) The explicit equation for chronic curve is $(X-X_0 / a)^2 + (Y-Y_0 / b)^2 = 1$. The constraint of the parameters that guarantees that the model will be an ellipse is $b^2 - 4ac < 0$.

f) Algebraic distance of ellipse $\approx q_i \approx L^TX_i \approx (d_i / d_i + r_i)$. Where d_i is distance of point beyond the ellipse and r_i is the radius. Points closer to the shorter axis of the ellipse have more effect on the fitting.

g) The objective function that needs to be minimized when fitting an ellipse using geometric distance is $d|p,f| = |P - X^*|$. which is the shortest distance between the point to arbitrary implicit curve. The additional challenge is to find the point on curve which is closest to the point.

h) The objective function for active contours is $E[\varnothing(s)]=\int(\alpha(s).E_{cont} + \beta(s).E_{curve} + \gamma(s).E_{img})ds$. Component Econt is continuity energy and it defines the gap between energy points, Ecurve is curvature energy and is used to handle curves in the contours and Eimg is image energy, it is used to adjust how close or further the contour needs to be from image borders. While α(s), β(s) and γ(s) are coefficients of the different energy items.

i) Continuity is the magnitude of derivative, Econt = $|d\varnothing / ds|^2 = |P_{i+1} - P_i|^2$. Curve energy is the magnitude of double derivative, Ecurve = $|d^2\varnothing / ds^2|^2 = |P_{i+1} - 2P_i + P_{i-1}|^2$.

j) The continuity of active contours may be relaxed by setting β(s) to 0 for the specific coordinates.

4. Model fitting 2:

a) Correlation matrix = $\begin{bmatrix} 10 & 14 \\ 14 & 20 \end{bmatrix}$

b) The implicit line equation is X.cos(45)+Y.sin(45) = d. Normalized normal is (1/√2,1/√2).

c) Y = -2.5

d) Vote d=1.

e) S = $\begin{bmatrix} 10 & 14 & 4 \\ 14 & 20 & 6 \\ 4 & 6 & 2 \end{bmatrix}$

f) The point p is at a distance of 0.5 from the implicit curve.

g) The point p is at an approximate distance of 0.5 from the implicit curve.

h) Continuity energy at p_2 is 2. Curvature energy at p_2 is 0.

i) B has to be set to 0 for high curvature point.

5. Robust estimation:

a) Outliers are points that are too far from the expected model outcome. The fundamental problem with outliers is that they are extreme and so they adversely affect the fitting and so the outcome varies i.e. it is inaccurate.

b) Objective function for robust estimation is $E(\theta) = \sum \theta_\sigma(d(x_i;\theta))$. In standard least square method the point which is furthest from the fitted model is eliminated, but the model in this case is already fitted and wrong. In robust, the model is fitted only with points which are inliers.

c) Geman-McClure function for robust estimation is $\nabla E(\theta) = \sum 2d\sigma^2 / (d^2+\sigma^2)^2$. We can start with a large σ and decrease as converging at each iteration.

d) The value of Geman-McClure estimator is 0.5

e) The principal of RANSAC is to perform multiple small experiments then choose the best results. Small sets are used with the hope that at least one set will not have outliers.

f) Parameters of RANSAC are ‘n’ number of points drawn at each evaluation, ‘d’ minimum number of points needed to estimate model, ‘k’ no of trials and ‘t’ distance threshold to identify inliers.
$K = \log(1-P) / \log(1-w^n)$ where P is probability of at-least one experiment is all inliers & w is probability that a point is an inlier.

g) Number of experiments should be 2.

6. Segmentation and recognition:

a) The objective of image segmentation is to identify objects in an image. It may simply do this by identifying the objects as foreground and rest as background. There are more complex methods to segment an image.

b) In an agglomerative we start with each pixel in a separate cluster then merge clusters with small distance i.e. similarity until cluster are not satisfactory. In divisive segmentation we start with all pixels in one cluster and split based on distance between them until clusters are not satisfactory.

c) In K-means algorithm we select K which is initial number of guessed clusters. Then repeat the process of assigning a cluster to the pixel as a label w.r.t the nearest cluster mean and average the new cluster mean until the cluster means stop changing.

d) In graph cuts we remove links with low similarity to create disconnected sub-graphs. This is not desirable as each pixel will end up being one sub-graph, so a normalized cut needs to be used and is necessary.

e) The similarity matrix will be a symmetric matrix of 100x100 dimension. The weighted degree matrix will be a diagonal matrix of size 100x100 and the Laplacian will be a symmetric and positive semi-definite matrix of size 100x100.

f) We can compute the eigenvector of Laplacian and stack a certain number of smallest vectors (say k) with non-zero values. So we get a n x k matrix. Then we take a section of this stack, i.e. k_i for all the n and cluster that with K means to get the cuts.

g) The optimization problem that has to be solved to determine a minimum normalized cut is to take into account the produced cluster at each node.

h) The solution to minimum normalized cut when using continuous variables is $\min_y (y^t(D-W)y / y^tDy)$ such that $y^tD1 = 0$.

i) In eigenfaces the image is mapped to a lower dimensional vectors using PCA then the similarity to templates in lower dimensional space is measured. This reduces the sensitivity to small variations in the image. Limitation is that the classification is in lower dimensions and so not as accurate.

j) In bag-of-words the features are extracted from an image using feature detection like HOG or SIFT. Then the features are clustered to create a codebook or dictionary of these features. Then we compute the distribution of code words in each class and then classify using the distribution of these code words. The limitation is that the image may identify objects which are perceptually make no sense.