1. Explain, with the aid of a diagram, how k-means clustering works in 2D.

Ans: 1. Randomly initialize the cluster centers.

2. For each point P, find the closest center, put P into cluster i.

3. Compute mean of points for each cluster i.

4. If center has changed, repeat step2.

2. Motivate multi-view stereo and explain how they work to resolve fundamental

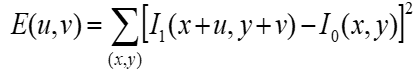
problems in stereo matching.

Ans: Fundamental problems: occlusion, ambiguity in matching

3. The SSD is typically used to do stereo matching. Write the equation for

SSD, and proceed to derive the expression for SSSD used in multibase stereo.

Ans:



5. What is the barber pole illusion? Explain its cause in the context of motion estimation.

Ans: Individual motion-sensitive neurons in the [visual system](https://en.wikipedia.org/wiki/Visual_system) have only limited information, as they see only a small portion of the visual field (a situation referred to as the ["aperture problem"](https://en.wikipedia.org/wiki/Motion_perception#The_aperture_problem)). In the absence of additional information, the visual system prefers the slowest possible motion: i.e., motion orthogonal to the moving line.

6. Relate the following concepts in two short sentences: homography, image rectification, and texture mapping.

Ans: In order to do texture mapping, we need to do image rectification with the help of homography transformation.

7. Please outline the algorithm for generalized Hough transform and explain it step by step.

Ans: HW1 Q2

8. Explain the voxel coloring algorithm in detail in the following context: color constraint, viewpoint constraint, depth ordering constraint, and outline the algorithm.

Ans: Color – shape from silhouette

Viewpoint – choose voxel; project and correlate; color if consistent.

Depth ordering – visit occluders first

9. Contrast forward mapping and backward mapping by illustrative examples.

Ans: Global visibility problem?

10. State no less than one advantage and one disadvantage on using large and small baseline, respectively, in stereo matching.

Ans: Large: difficult search problem

Small: large depth error

11. Explain the multi-baseline stereo algorithm, in terms of the following:

(a) reference view,

(b) baseline arrangement, and

(c) multi-baseline matching.

Ans: (a) Choose a reference view

(b) Baseline arrangement

(c) Replace two view-SSD with SSD over all baselines

12. Discuss the computability and complexity of photo-consistent scenes in multi-view stereo, given a quantized volume consisting of N^3 voxels, and C colors.

Ans: All scenes: C^(N^3)

13. In two sentences, describe the reconstruction algorithm from silhouettes, given a set of calibrated silhouette images. What are the major advantages of this algorithm?

Ans: (1) Project each silhouette

(2) Intersect projected volumes

(3) Color voxel back if on silhouette in every image

Advantage: easy to implement, fast; accelerated via octrees

14. Consider the reconstruction from silhouette algorithm that colors voxel by performing volume intersection in the 3D space. What do we get in the limit? What is the time complexity of this algorithm, given M images and N^3 voxels?

Ans: Reconstruction contains the true scene

O(MN3)

15. State no less than 3 advantages and disadvantages (combined) of volume intersection (voxel coloring given silhouettes).

Ans: Pros: Easy to implement, fast; accelerated via octrees

Cons: No concavities; reconstruction is not photo-consistent; requires identification of silhouettes

16. Describe by listing the steps in voxel coloring algorithm proposed by Seitz and Dyer that computes a photo-consistent reconstruction, when viewpoints are constrained. What is the crucial property of the camera arrangement to ensure the correctness of the algorithm?

Ans: choose voxel project and correlate color if consistent

Depth order is the same for all input views

17. State and describe no less than three camera arrangements so that scene depth ordering can be computed for voxel coloring algorithm proposed by Seitz and Dyer.

Ans: Overhead inward-facing camera moved 360 degrees around an object

Rig of outward-facing cameras distributed around a sphere.

18. Describe by itemizing the main steps of the space carving algorithm discussed in class. What are the two lemmas we covered in class to ensure the convergence of the algorithm?

Ans: (1) Initialize to a volume V containing the true scene

(2) Choose a voxel on the current surface

(3) Project to visible input images

(4) Carve if not photo-consistent

(5) Repeat until convergence

Visibility lemma

Non-photo-consistency lemma

19. Explain the visibility lemma on the white board.

Ans: Let p be a point on surface of V, Surf(V), and let Visv(p) be the collection of input images in which V does not occlude p. If V’, a subset of V, is a shape that also has p on its surface, Visv(p) is a subset of Visv’(p).

20. Explain scaled orthographic projection, and state no less than two conditions scaled orthographic projection can be used to approximate perspective projection.

Ans: Orthographic first, then perspective

1. Close to the optical axis
2. Dimensions are small

21. In structure-from-motion for two views, for n points there are 3n unknowns, and 5 unknowns in motion for two views. How many correspondences are required? Write down the inequality.

Ans: Each point correspondence gives 4 measurements

For n points expect a solution if 4n >= 3n+5 n>=5

22. Give at least one example on the following image processing functions.

(a) Range transformation,

(b) Domain transformation.

Ans: (a) Mean filter

(b) Image warping

23. What is a Gaussian Pyramid? What are the two main steps to construct a Gaussian Pyramid?

Ans: (1) Filter

(2) Subsample

Represent NxN image as a “pyramid” of 1x1, 2x2, 4x4 … images

24. What is a linear shift-invariant (LSI) filter? Is convolution a LSI filter?

Ans: linear and shift-invariant

Yes

25. Distinguish no less than two types of image noises and the corresponding effective method to remove them.

Ans: Salt and pepper & gaussian & impulse

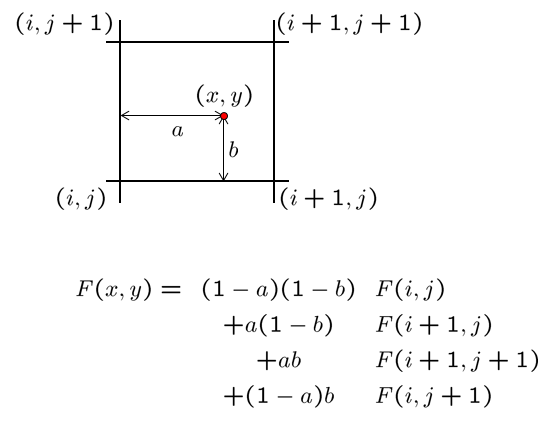
Median filter & Mean filter & Mean filter

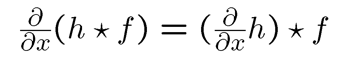
26. Will alias occur if the sampling rate is higher than the highest frequency of the signal? What is the Nyquist frequency?

Ans: Probably

Twice times the highest frequency

27. Draw the 2D shape of the bilinear interpolation function and write down the expression. Explain the expression.



28. State the derivative theorem of convolution, and one advantage it can bring. 

Reduce one calculation.

29. In thinning, will the result that produced by non-maximum suppression the same as thresholding? Why is it necessary to perform thinning before edge following?

Ans: Different

Or the edge might be thick

30. Name no less than two origins of edges we observe in images and explain what they are.

Ans: Surface normal disparity

Depth disparity

Surface color disparity

Illumination disparity

31. What is the effect of sigma in the Gaussian kernel on the result of edge detection.

Ans: Large sigma detects large scale edges

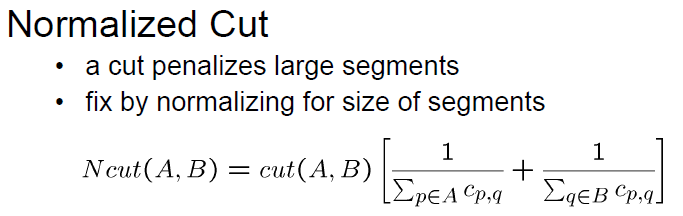
Small sigma detects fine features

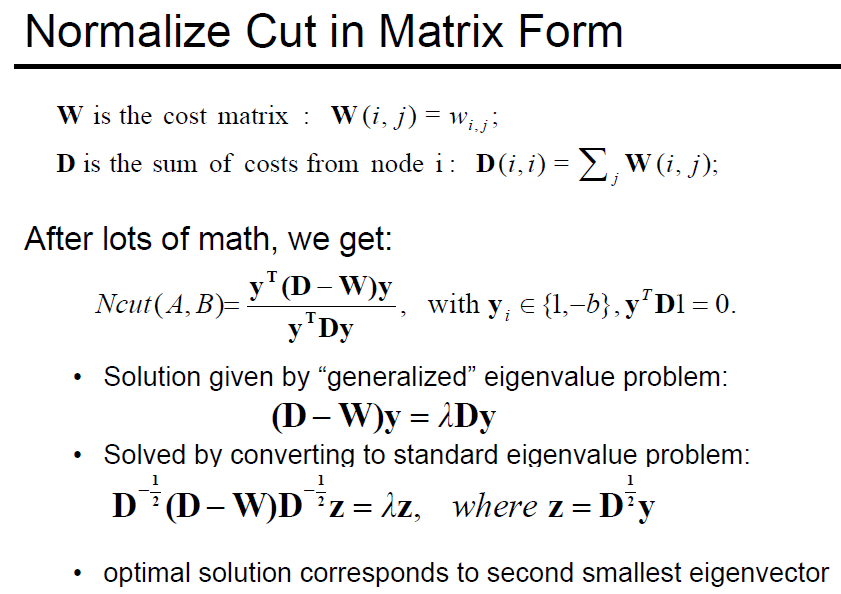
32. Name no less than two Gestalt laws of perceptual organization. Explain them with illustrative examples.

Ans: Proximity, similarity, continuation, closure, common fate

33. Normalized cuts can be used to perform image segmentation. To begin, this graph theoretic approach takes as input a fully connected graph. Outline the rest of the algorithm.

Ans: The cost for each link measures dissimilarity (difference in color and position). Then break the graph into segments:





34. Outline histogram-based segmentation. In your opinion, what are the advantages and disadvantages?

Ans: Advantages: fast, intuitive

Disadvantages: only work for morphological graph, not accurate

35. Name no less than 2 morphological operators and illustrate on the whiteboard using a 2D binary image how they work.

Ans: dilation & erosion & opening & closing

36. Suggest and explain no less than two extensions of Hough transform (except the generalized Hough transform for shape detection).

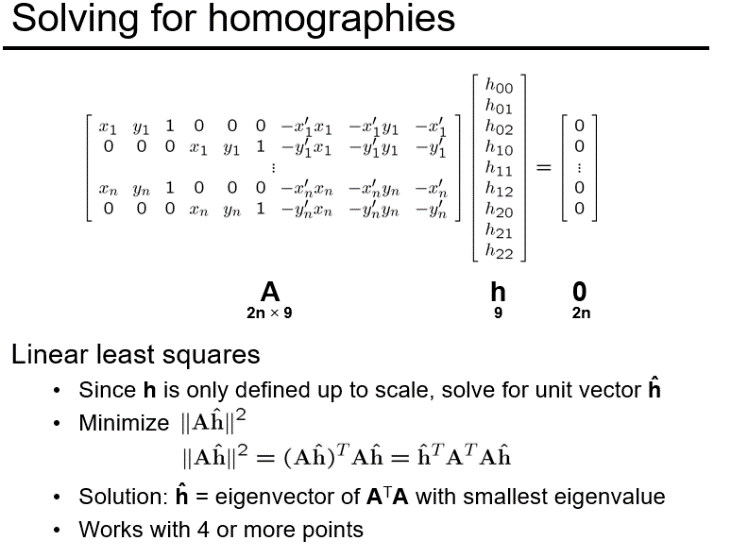
Ans: 1. Use the image gradient

2. Give more votes for stronger edges

3. Change the sampling of (d, theta) to give more/less resolution

4. Circles, squares or any other shape

37. Write out the derivation of homography and show that at least 4 correspondences are needed to determine the homography matrix.



38. What is the geometric intuition of homogeneous coordinates? Draw it on the white board.

Ans: A point in the image is a ray in projective space

A line is a plane of rays through origin

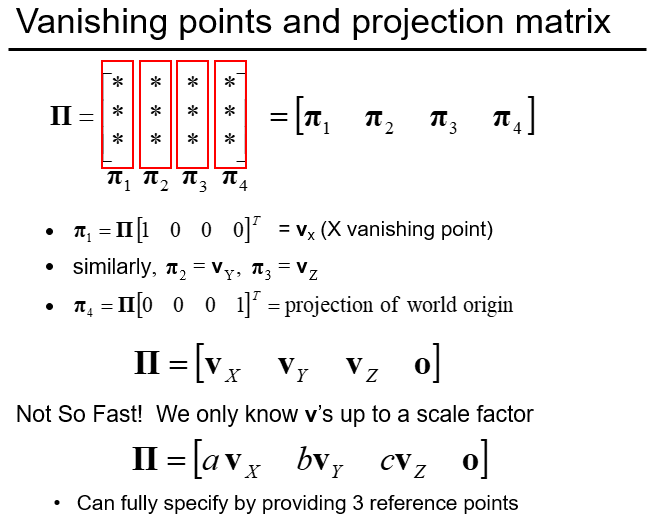
39. Derive the expression for vanishing points on the 2D image plane.

Ans: (P1 X P2) X (P3 X P4)

40. Define cross ratio.

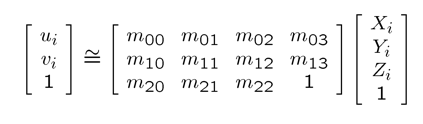
Ans: A projective invariant – something that does not change under projective transformations (including perspective projection)

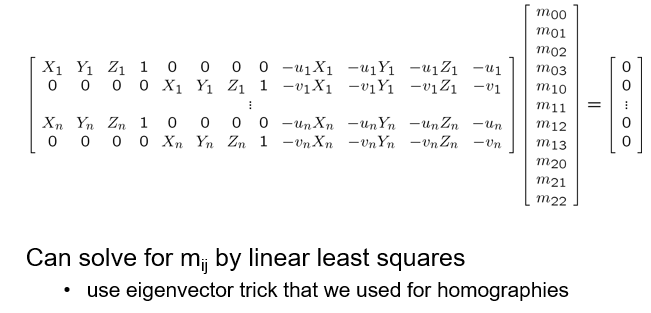
41. Derive the 3x4 projection matrix, given vx, vy, vz, o (the principal vanishing points) and the world origin.

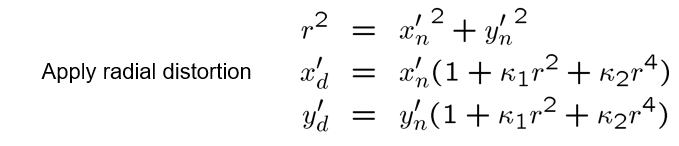


42. Write and expand the direct linear calibration and show that the minimum number of 5-tuples to compute the projection matrix is 6.

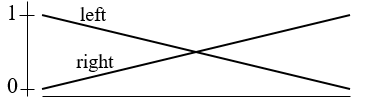
Ans: up to scale



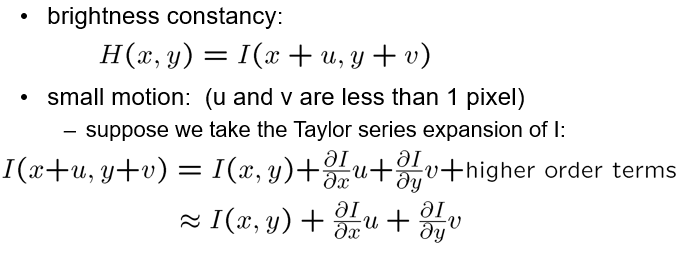


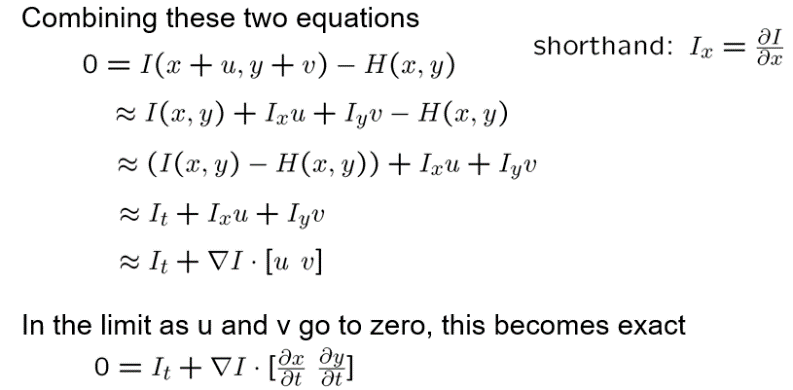
43. Define radial distortion and write down the pertinent expressions described in class.

44. Plot the following blending function. C = alpha F + (1-alpha) B

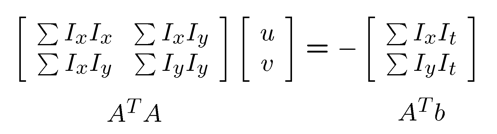


45. Define color constancy and small motion using equations. Then, derive the optical flow equation. How many unknowns to solve for?

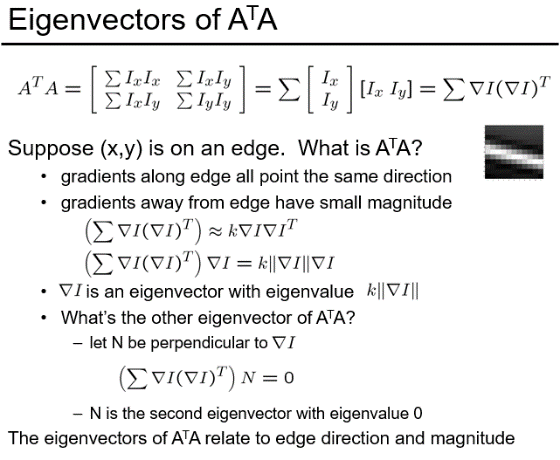




46. Derive the Lukas-Kanade equation from the optical flow equation.



47. [Give Lukas-Kanade equation]. Derive the eigenvectors and eigenvalues of A^TA. State the significance of eigenvectors.



48. Describe the three cases represented by the eigensystem iof \nabla I \nabla I^T.

Ans: Edge – gradients are all the same or close to 0, large nambda\_1, small nambda\_2

Low textured region – gradients have small magnitude, small nambda\_1, small nambda\_2

High textured region – gradients are different, large nambda\_1, large nambda\_2

49. State the three basic steps in stereo reconstruction. Draw a figure to help illustrate.

Ans: Calibrate cameras

Rectify images

Compute disparity

Estimate depth

50. Give three points that define an epipolar plane. Define an epipolar line using the epipolar plane.

Ans: Two center of projection + object point

Epipolar line = epipolar plane intersect with images

51. What is an epipole? What is its significance?

Ans: Intersection between the image plane and baseline.

Can have infinite image coordinates? If rotate epipolar planes along the baseline, all epipolar lines intersect at epipole.

52. Let F be the 3x3 fundamental matrix relating I\_1 and I\_2. Let p\_1 \in I\_1 and p\_2 \in I\_2 are two corresponding points. Write out the expression of the two epipolar lines.

Ans: Fp\_1 & FTp\_2

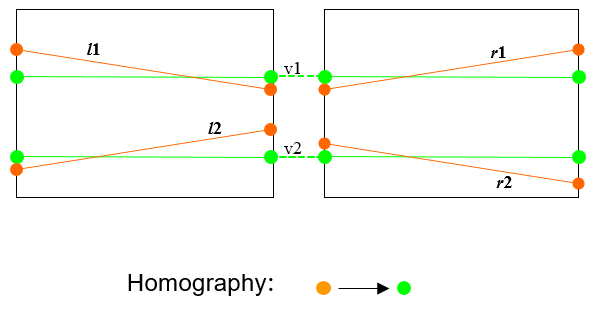
53. How is image reprojection related to image rectification? Why image rectification is desirable for stereo reconstruction?

Ans: Stereo image rectification = reproject image planes onto common plane parallel to line between optical centers

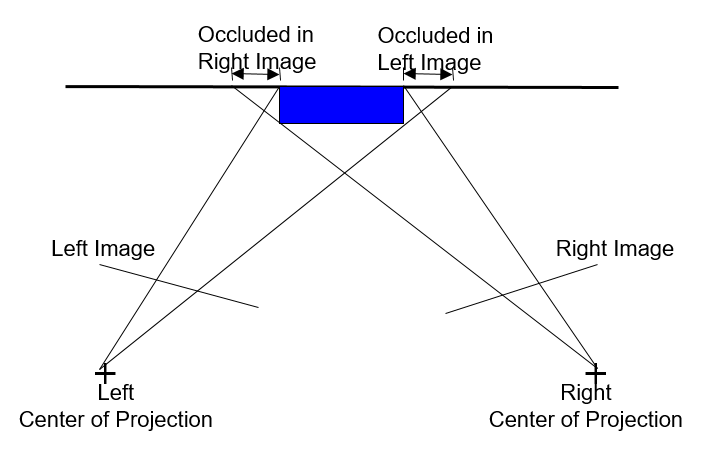
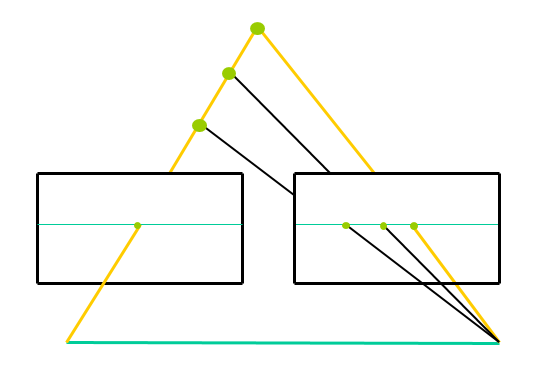
Usually the image planes of a given stereo pair are not parallel, after rectification, the image planes become coplanar and parallel to the baseline, conjugate epipolar lines become corresponding scanlines.

54. Describe a simple image rectification algorithm that is based on homography described in class.

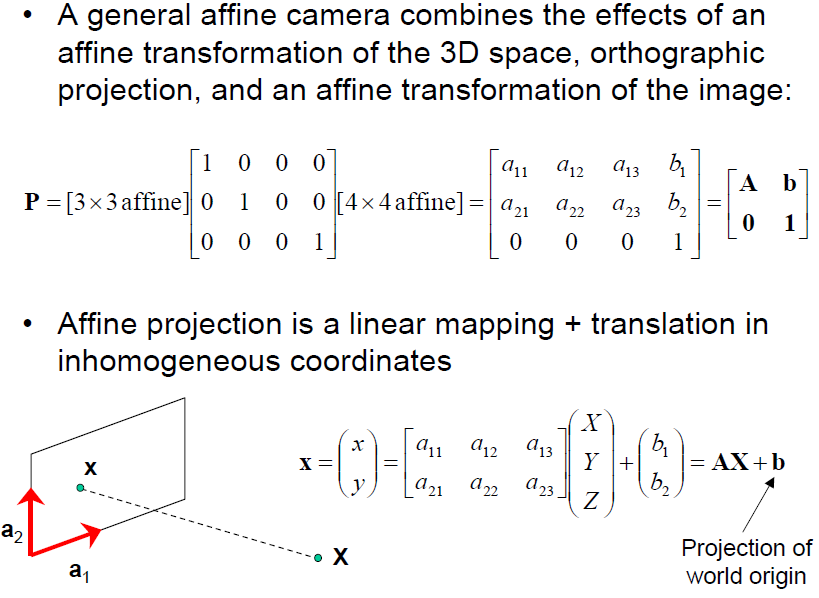
Ans:



55. Describe two fundamental problems in stereo matching, and how they are related to the baseline chosen (if applicable).



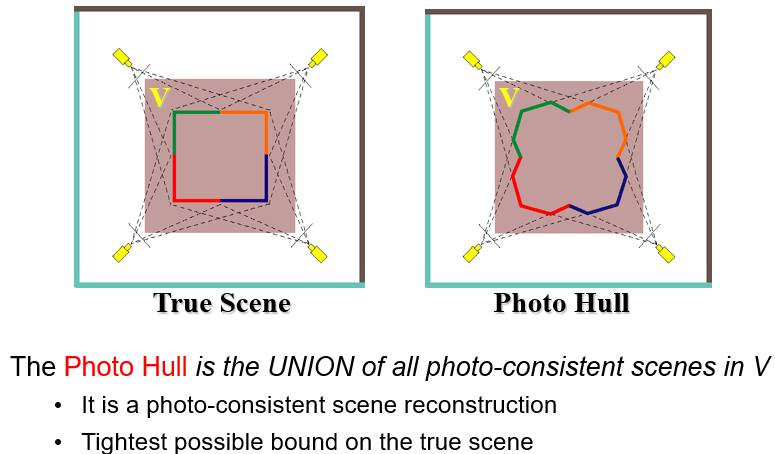
56. What is an affine camera. Can you write down an affine camera using inhomogeneous coordinates described in class?



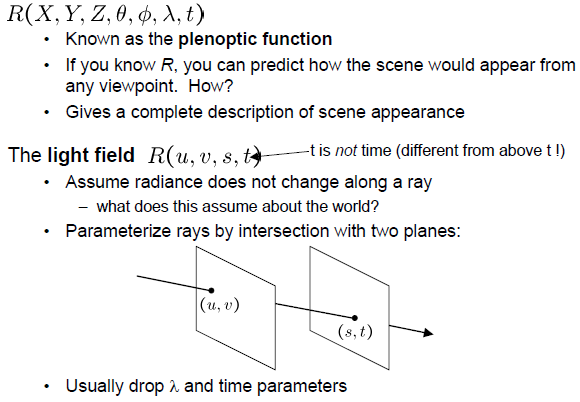
57. For affine structure from motion, given m images and n fixed 3D points, how many structure and motion unknows to solve? For two views, how many point correspondences are required?

Ans: 2mn >= 8m+3n-12 4points

58. Consider a plane half colored green and half colored red (draw the figure on white board), what is the output by the space carving algorithm?



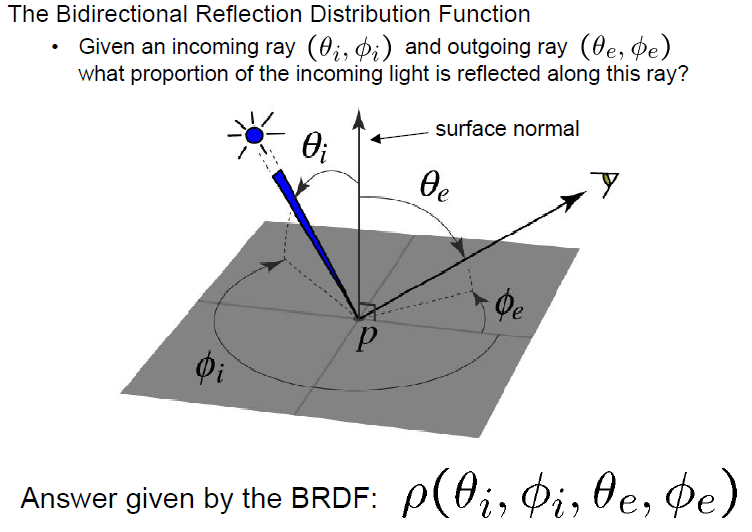
59. Describe the seven parameters of a plenoptic function. What is a light field?



60. Rods and cones act as filters on the visible spectrum. Describe in one or two sentences how the get the output of such filters.

Ans: Multiply its response curve by the spectrum, integrate over all wavelengths.

61. Define BRDF using a 2D diagram.

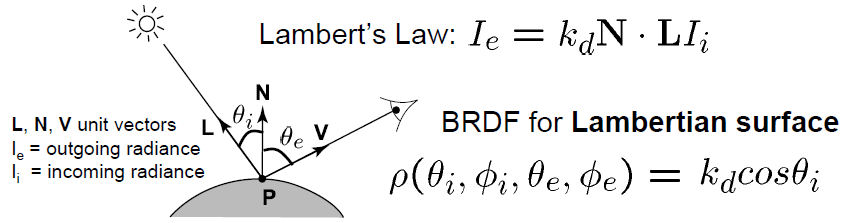


62. In perfect diffuse reflection, how is the viewed brightness related to the viewing angle? How is the brightness related to the direction of illumination?

Ans: Viewed brightness does not depend on the viewing angle

Brightness does depend on the direction of illumination

63. What is the BRDF for a Lambertian surface. Does it depend on all the input parameters?



64. How do you design a probabilistic skin detector, given P(skin|R)?

Ans: Choose interpretation of highest probability

65. In order to learn the posterior probability distribution function (posterior pdf) or P(skin|R), we calculate P(R|skin). How do you find P(R|skin)?

Ans: A histogram over the pixels in the training images

66. Define Bayses law using the P(skin|R) by an equation, and explain the all the terms on the right-hand side. You may skip the "normalization term".

Ans: P(skin|R) = (P(R|skin) \* P(skin)) / P(R)

From left to right: posterior, likelihood, prior, normalization term

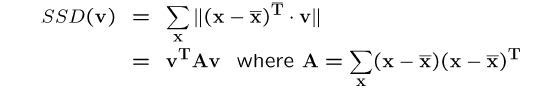
67. What is MAP (maximum a posterior) estimation? What is ML (maximum likelihood) estimation? Which one is easier to find? Under what condition can ML be used as MAP?

Ans: MAP - chooses the label that maximize the posterior

ML – choose the label that maximize the likelihood

When prior is uniform: P(skin) = P(~skin) = 0.5

68. What is principal component analysis (PCA) for estimation in N-dimension? Define it using an equation. Explain in one sentence how PCA is used to reduce the dimensionality of the parameter space, which can sometimes be prohibitive.



Eigenvectors of A define a new coordinate system, which can be used to form a “hyper-plane”

69. Consider the eigenface algorithm. Suppose our face database consists of face images of dimension nxn, and that we reduce the dimensionality by PCA by choosing the top 10 eigenvectors which serves as the basis for face representation. What is the dimensionality of the eigenvectors? What is the dimensionality of the face representation?

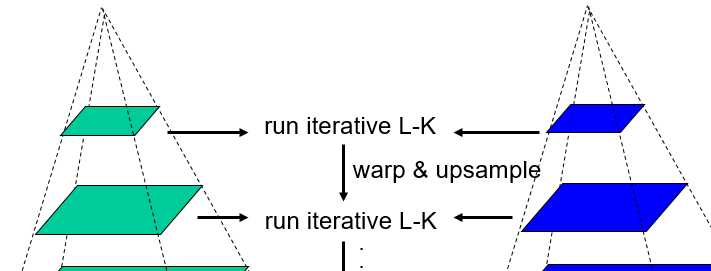
Ans: nxn 1x10

70. State the difference between convolution and cross correlation.

Ans: Convolution flips the filter horizontally and vertically before applying it.

71. Explain (a) the motivation why we need the coarse-to-fine motion estimation, and (b) how the solution produced in a lower resolution helps to compute the solution in a higher resolution, given that we use iterative optimization method (e.g. Newton's root finding) for solving the objective functional.

Ans: (a) Because motion not small enough

(b) 

72. What is a first-order Markov Chain and nth-order Markov Chain? Express them in terms of random variables x\_t, x\_{t-1}, ...

Ans: first: p(xt|xt-1)

N-th: p(xt|xt-1, …,xt-N)

73. Briefly describe the two steps in text synthesis that makes use of Markov chain approach.

Ans: (1) Build probability histogram, find all blocks of N consecutive words in training set, compute probability of occurrence p(xt|xt-1, …,xt-(n-1))

(2) Given words x1, x2, …, xk-1, compute xk

74. What is a first-order Markov Random Field in the discrete 2D space? Illustrate by drawing a diagram on the whiteboard.

Ans: P(X|A, B, C, D)

Four neighbors

75. What will be the effect of the window size in non-parametric texture synthesis?

Ans: Windows size controls regularity

76. For patch-based texture synthesis, suggest at least two methods commonly used by vision researchers to reduce the seams in the overlapping area.

Ans: Minimal error boundary cut

Neighboring blocks constrained by overlap

77. In patch-based texture synthesis ("copy and paste"), name two guidelines we described in class that concerns with the fill order.

Ans: Choose pixels that have more neighbors filled

Choose pixels that are continuous of lines/curves/edges

78. The stereo reconstruction can be formulated into one of graph cuts.

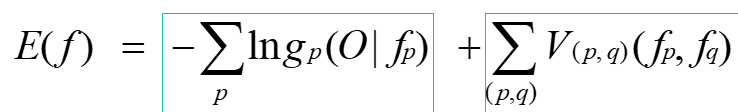
(a) How do you map to the vertices and links of the underlying graphs?

(b) What is the label set?

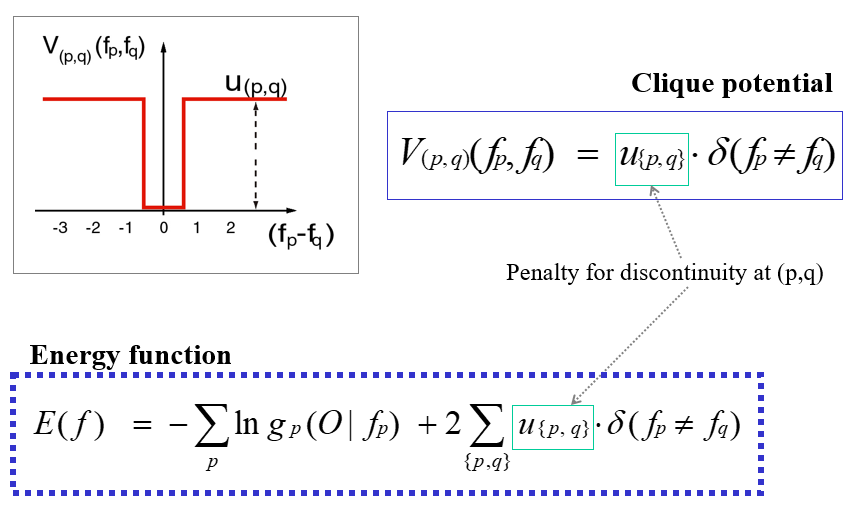
Ans: (a) pixel + label = vertices; pixel-pixel link and pixel label link = links

(b) All possible disparity

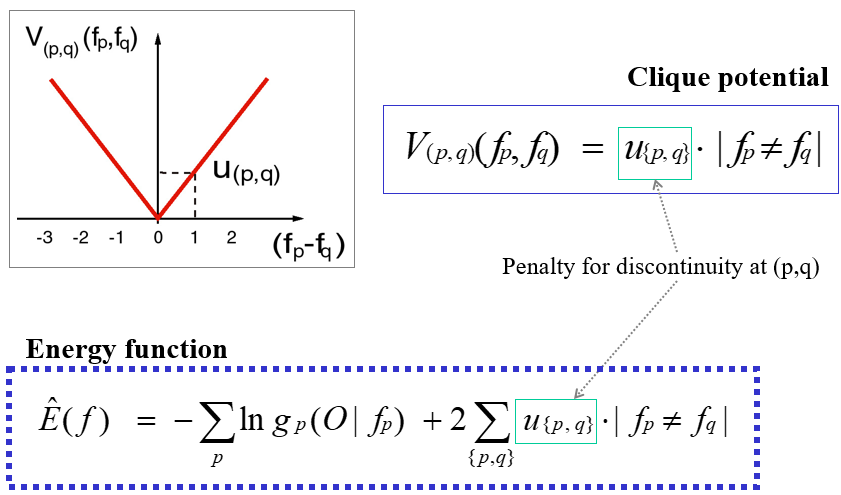
79. Write down the MRF energy minimization function E(f) for disparity estimation, where f\_p is the configuration at a pixel p, O is the observed data, and V is the compatibility function.



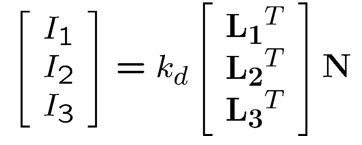
80. Draw the generalized Potts model on the whiteboard and explain the intuition behind its design.



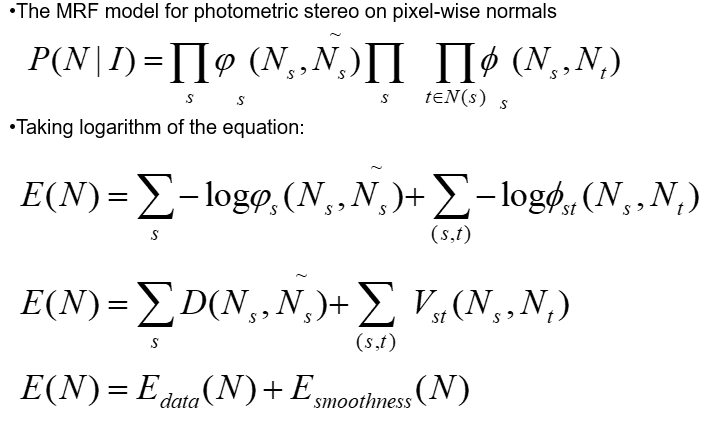
81. Draw the linear clique potential model on the whiteboard and explain the intuition behind its design.



82. Given three images I\_1, I\_2, and I\_3 with corresponding light sources L\_1, L\_2, L\_3, write down the equations for photometric reconstruction in a matrix equation.



83. Write down the equation of MRF model for photometric stereo for normal estimation and explain each term.

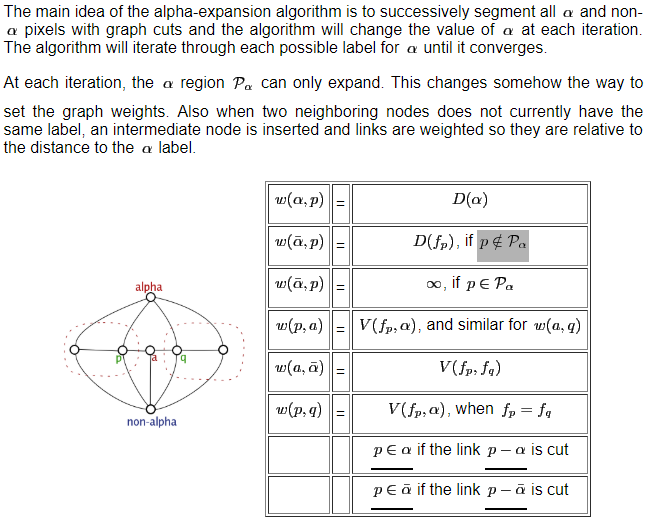


84. Consider the first-order MRF. On what condition the alpha expansion can be used and why it is a faster algorithm?

Ans: If the energy function is regular

It is linear.

85. Please outline the alpha-expansion algorithm.



86. Define the shape from shading problem using the Lambertian model. Name two of the three constraints that you can use to obtain a good solution.

Ans: Directly measure angle between normal and light source

1. Assume a few of the normal are known
2. Constraints on neighboring normals – “integrability”
3. Smoothness

87. Outline the "multiway cut algorithm" that makes use of the maxflow algorithm in each iteration for stereo reconstruction.

Ans: Alpha-Expansion

88. Outline the two steps in synthesizing one pixel in non-parametric texture synthesis:

a) find matching neighborhood

b) how to synthesize the pixel

Ans: a) Consider only pixels in the neighborhood that are already filled in, find all windows in the image that match the neighborhood.

b) pick one matching window a random, assign x to be the center of that window

89. There are two steps in synthesizing one pixel in non-parametric texture synthesis, namely,

a) find matching neighborhood

b) how to synthesize the pixel

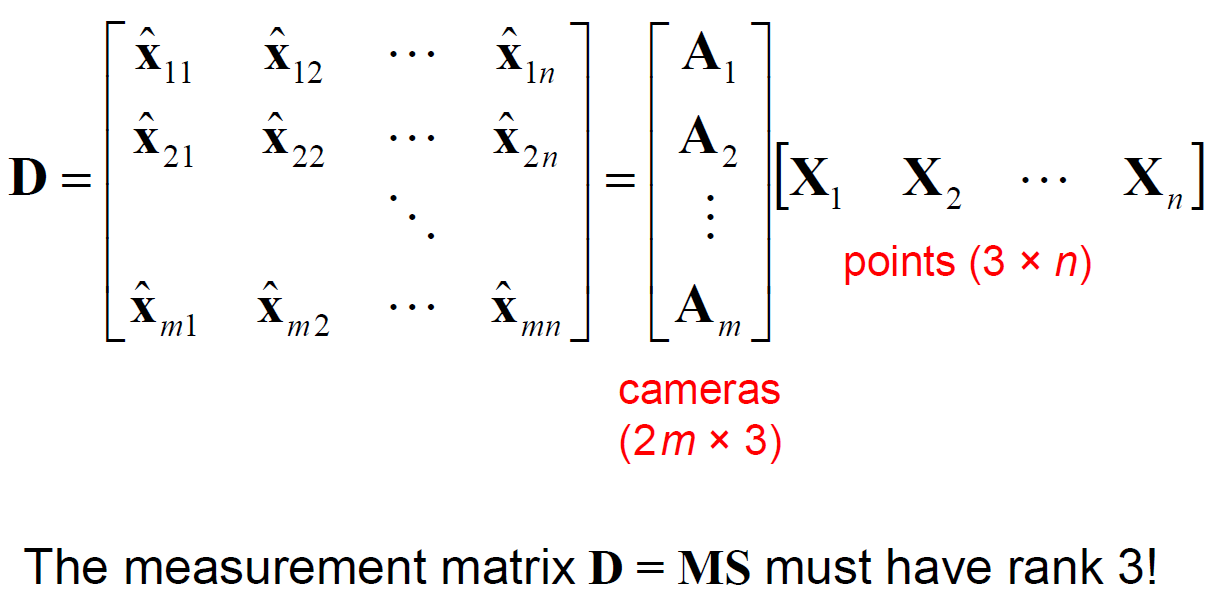
How will you modify the above into a block-based texture synthesis?

What additional consideration needs to be used in order to avoid noticeable artifacts such as seams?

Ans: Choose pixels that have more neighbors filled

Choose pixels that are continuous pf lines/curves/edges

90. Consider the 2m x n data matrix, where 2m is the number of cameras and n is the number of points. Factorize it into the camera and points matrix. What is the property on the measurement matrix regarding to its rank?



91. In the general structure and motion problem, given n matching image points and m views,

(a) For each camera, how many parameters to be estimated (if translation magnitude can also be estimated)?

(b) For each 3D point, how many parameters to be estimated?

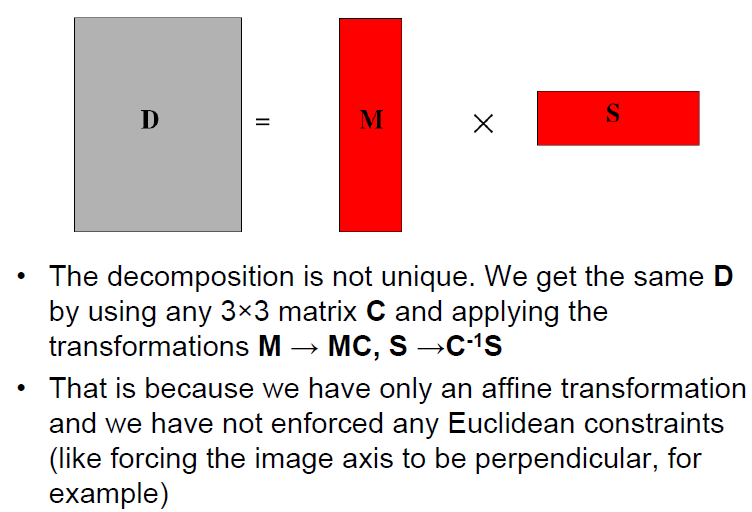
(c) What is the total number of parameters to be estimated?

Ans: (a) 6

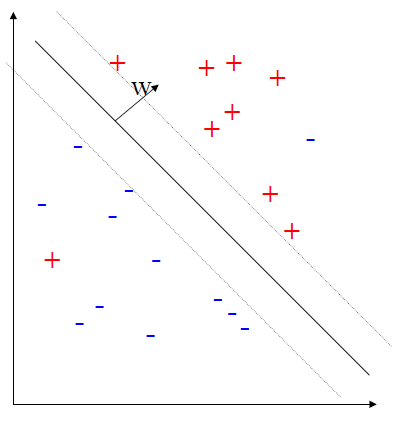
(b) 3

(c) 6m+3n

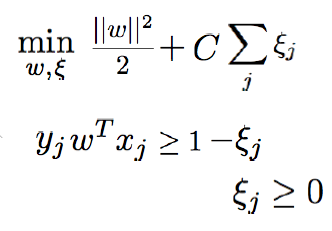
92. In Tomasi-Kanade SfM, single value decomposition is used to factorize the data matrix. Enforcing the rank-3 constraint and eliminating the affine ambiguity are done after the factorization. Describe how these two post-processing are done.



93. What is linear maximum-margin hyperplane in SVM for solving the classification problem: given a training dataset of n points of the form (\vec{x}\_1, y\_1),\, \ldots ,\, (\vec{x}\_n, y\_n) where the y\_i are either 1 or -1, each indicating the class to which the point \vec{x}\_i belongs.



94. Write down the objective function for learning maximum-margin SVM.



95. Explain HOG in feature extraction/encoding.

Ans: Histogram of gradients

96. Explain sliding window at multiple scales for object detection.

Ans: Project2

97. How do you adapt a single scale detector to detect object at multiple scales?

Ans: Adjust size of images.

98. Let F(x, theta) be the decision function and y(x) be the true label. Define the hinge loss of a linear SVM.

Ans: L(y(x),F(x, theta)) = max(0,1-y(x)F(x, theta))

99. What is overfitting in SVM learning? Why do we prefer a simple model?

Ans: Overfitting refers to a model that models the training data too well, it happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data.

To avoid overfitting?

100. A dataset is typically split into three in SVM learning. What are they and what is the role of each split set?

Ans: Training set

Validation set

Test set