Computer Vision

Q) What are 2D motion models? Explain how they can be used for image stitching?

A:→ 2D motion models are models that are used to establish mathematical relationships ~~form~~ one image to another for registering and aligning.

→ Image stitching algorithms for aligning images and stitching them into seamless photo-mosaics are among the oldest technique.

→ They are used to create high resolution photo-mosaics to produce today's digital maps and satellite photos.

→ Also used to create wide angle panaromas.

→ Image stitching comes under photogrammatery.

→ Image stiching acts as a base for bundle adjustments.

→ Used to remove 'ghosts' and parallax

→ Image stitching is faster than older techniques like pixel to pixel matching, sparse set of features

→ Steps:

proper mathematical model

↓

pixel to pixel matching using gradient

↓

finally, Aligning images

Q2) Explain the basic steps involved in choosing a composting surface.

→ Composting is process of getting finally stitched mosaic image.

→ This involves selecting a final composting surface (flat, cylindrical, spherical etc) and reference image.

→ The final step is to how optimally blend these pixels to minimize visible seams blur ~~goa~~ ghosting.

Choosing a composting surface:

- The first choice is how to represent image.
- If only few pictures are there, select a reference image from one of those.
- For few images it is sometimes called as flat panorama
- For larger view we select cylindrical / Spherical projection

⇒ View selection:

→ which part of the scene will be centered in the final view.

→ In case of smaller image → one of those pics acts as reference

→ In case of larger spherical / cylindrical ~~coordinates~~ projection taken.

Date
Page

⇒ coordinate Transformation: next step is
⊙ mapping of coordinates between input and output pixels
⊙ If final composting surface is flat, ~~no~~ i/p image ~~...~~ have no radial distortion
⊙ If final composting surface is cylindrical or spherical, then i/p has radial distortion

⇒ Sampling issues:
sometimes if all the above said process yield the correct pixel address there may be a problem of low resolution.
→ we can use MIP mapping to correct that in order to get high quality images.

(3) With neat figure explain the projection from
a) 3D to cylindrical coordinate
b) 3D to spherical coordinate.

a) 3D to cylindrical coordinate:
3D coordinates are given by (θ, h) :-

$$(\sin\theta, h, \cos\theta) \propto (x, y, f)$$

$$x' = s\theta = s\tan^{-1}\frac{x}{f}$$

$$y' = sh = s\frac{y}{\sqrt{x^2+f^2}}$$

The inverse of this mapping equations is given by

$$x = f \tan\theta = f \tan\frac{x'}{s}.$$

$$y = f\frac{y'}{s} \sec\frac{x'}{s}.$$

b) 3D to spherical coordinates

→ we have 2 angles (θ, ϕ).
→ coordinates are given by $(\sin\theta\cos\phi, \sin\phi, \cos\theta\cos\phi)$
→ Mainly ~~hel~~ helpful if panaroma includes a full sphere or hemisphere views
→ spherical coordinate system are mainly used by professional panaromic photographers.

Coordinates:

$$(\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi) = s(x, y, z)$$

$$x' = s\phi\cos\theta = s\frac{x}{r}\tan^{-1}\frac{r}{z}$$

$$y' = s\phi\sin\theta = s\frac{y}{r}\tan^{-1}\frac{r}{z}$$

(4) a) Explain the process of recognising panoramas?

b) What are various difficulties faced while recognizing panoramas?

4) How do panaromas are recognised? Explain various methods ~~in~~ for recognising panaromas

→ Panaromas are created when image is captured in elongated view.

→ Basically it is stitching multiple images together.

→ A 360° photo is a controllable panoramic image that surrounds the original point from which shot was taken.

The various methods

1) Bundle adjustment :-

2) Feature matching :

(i) corresponding features from 2 images.

(ii) One is source image. Other one is target image.

3) Image matching :

→ process of bringing two images geometrically into so that pixels of the 2 images correspond to same physical.

→ Harris corner

→ SIFT (Scale Invariant feature transform)

↓

detecting local features

(4) a) Explain the process of recognising panoramas?

b) What are various difficulties faced while recognizing panoramas?

(5) i) Write short notes on following 2D geometric primitives.

a) 2D points (2)

b) 2D lines (2)

(ii) Explain the basic set of 2D planar transformation with neat figure.

a) 2D points:

→ 2D points can be denoted by using pair of values $x = (x, y)$

$$x = \begin{bmatrix} x \\ y \end{bmatrix}$$

→ In case of homogeneous coordinates :-

$\tilde{x} = (\tilde{x}, \tilde{y}, \tilde{w}) \in P^2$.

↳ projective space.

b) 2D lines:

$l = (a, b, c)$

$l = ax + by + c = 0$.

Homogeneous coordinates:

$\tilde{x} = \tilde{l}_1 \times \tilde{l}_2$

$\tilde{l} = \tilde{x}_1 \times \tilde{x}_2$

(ii) 2D Transformations :

⊙ Translation : $x' = x + t$

$$x' = \begin{bmatrix} I & t \end{bmatrix} \bar{x}$$

⊙ Rotation + translation

$$x' = \begin{bmatrix} R & t \end{bmatrix} \bar{x}$$

⊙ scaled rotation :

$$x' = \begin{bmatrix} a & -b & t_x \\ b & a & t_y \end{bmatrix} \bar{x}$$

⊙ Affine :

$$x' = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix} \bar{x}$$

⊙ projective :

$$x' = \frac{h_{00}x + h_{01}y + h_{02}}{h_{20}x + h_{21}y + h_{22}}$$

$$y' = \frac{h_{10}x + h_{11}y + h_{12}}{h_{20}x + h_{21}y + h_{22}}.$$

(6) i) Explain the general image processing operator in detail.

→ The simplest kind of image processing transforms are point operators where each output pixel's value depends on only the corresponding input value.

→ They include brightness & contrast adjustments, color corrections and transformations.

→ In the image processing literature such operations are also called as point processes.

→ A pixel in context of CV, is the numerical value of the scalar or vector information at one point in a picture

→ pixel transform is to convert it from one domain to another.

(ii) Define the following :-

a) composting :-

→ composting is process of getting finally stitched mosaic image.

→ It involves selecting a final composting surface (flat, cylindrical, spherical etc)

b) Matting :

→ The process of extracting the object from the original image is often called matting.

→ The actual process of matting is recovering

the foreground, background and alpha
matte values from one or more images.

(7) i) What is Bidirectional reflectance Distribution function (BDRF)? (4)

(ii) With neat figure, discuss the components into which a typical BDRF is split into? (6).

(i) ·) BDRF → used for scattering of light.

→ It has 4 function namely

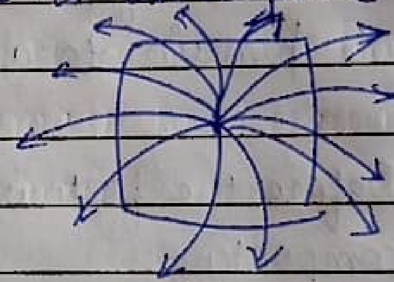
$$R(a, b, c, d, \lambda)$$

each has seperate wavelength in it.

→ It has reciprocal formula and that formula is Helmholtz formula

$$R(\hat{a}, \hat{b}, \hat{c}, \hat{d}, \lambda)$$

→ The scattering of light is isotrophic



→ It will not be in perfect direction, so it is iso-trophic.

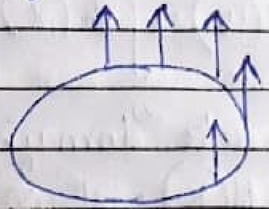
(ii) Basically BRDFs can often be split into their diffuse and specular components

→ Diffuse component:

The diffuse component scatters light uniformly in all directions mostly associate with shading.

→ It often imparts a strong body color

→ Since scattering is uniform in all directions BRDF is constant.



→ Specular reflection:

→ Specular reflection depends strongly on the direction of outgoing light.

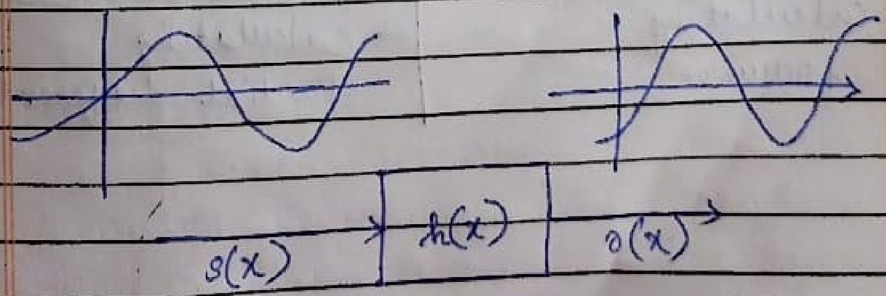
→ Incident light rays are reflected in direction that is rotated by 180° around the surface normal.

(8) (i) What is Fourier Transform? Discuss about the closed form equation of FT in continuous and discrete domain.

→ Fourier transforms are used to convert signals to frequency domain.

→ It can be ~~either~~ represented in either sum of sin or cos waves.

→ Fourier transformation → signals can be converted from time to frequency.



→ closed form equations for fourier transform exist both in continuous domain

$$H(\omega) = \int_{-\infty}^{\infty} h(x) e^{-j\omega x} dx$$

→ In discrete domain

$$H(k) = \frac{1}{N} \sum_{x=0}^{n-1} h(x) e^{-j\frac{2\pi k x}{N}}$$

Discrete FT takes $O(N^2)$ operations to evaluate.

(ii) Differentiate between discrete & fast FT

FFT	DFT
→ Faster version of FT	→ Discrete version of FT
→ All fast computations collectively called as FFT algorithm	→ It converts time domain to frequency domain
→ FFT is used in DFT	→ ~~Relation between~~ * Relation between time & freq.
Application :- → calculating fouriers	Application → Calculating Partial differentiation.

(9) a) What is need for projective reconstruction?

→ Projective reconstruction refers to the computation of the structure of Scene from images taken with uncalibrated cameras.

→ Basic constraint : $x^T E x_0 = 0$

→ The matrix is given as

$$\begin{bmatrix} u_0 & u_1 & e_1 \end{bmatrix} \begin{bmatrix} \sigma_0 & & \\ & \sigma_1 & \\ & & 0 \end{bmatrix} \begin{bmatrix} V_0^T \\ V_1^T \\ e_0^T \end{bmatrix}$$

→ projective reconstruction can be used for Homogeneous images

→ It is also called as plane plus parallax.

b) Briefly discuss why does essential matrix change into Fundamental matrix?

$$F = \begin{bmatrix} u_0 & u_1 & e_1 \end{bmatrix} \begin{bmatrix} \sigma_0 & & \\ & \sigma_1 & \\ & & 0 \end{bmatrix} \begin{bmatrix} V_0^T \\ V_1^T \\ e_0^T \end{bmatrix}$$

$$H = K_1^{-T} R K_0^{-1}$$

$$F = U \Sigma V^T$$

$$\tilde{H} = U R_{90}^T \Sigma V^T$$

$$P_0 = [I | 0]$$

(10) Explain Hierarchical motion estimation in detail.

→ It is a process ~~of~~ where, an image pyramid is constructed and a search over a smaller number of discrete pixels is performed.

→ The motion estimate from one level of the pyramid is then used to initialise a smaller local search for next level.

→ $I_k(x_j) \leftarrow I_k^{(l-1)}(2x_j)$

→ At coarest level, we search for best displacement u and that minimises the difference between images I_0 and I_1

→ One suitable vector is used to predict the displacement.

→ The search over displacements then repeated.

→ Alternatively, one of the images can be warped by the current motion estimate, in which case only small incremental motions need to be computed at the finer level.

(11) Bundle adjustments.
(12) Fourier based aligmment.

13 a) Explain snakes in detail. what is main drawback in using snakes. (6)
b) what is the basic solution to overcome the drawback while using snakes (4)

a) → The snakes in CV are also called as Active contours
→ It works like stretched elastic band being released
→ It is to identify and outline the target objects
→ Snakes are configured by use of spline

→ Used in medical imaging.
→ The main drawback of snakes is noise sensitivity

b) → we can use splines and shape priors.
→ aka B-Spline approximations.
→ The resulting B-Snake can be written as

$$F = BX$$

$$F = \begin{bmatrix} f^T(0) \\ \vdots \\ f^T(N) \end{bmatrix}, \quad B = \begin{bmatrix} B_0(s_0) & \cdots & B_K(s_0) \\ \vdots & \ddots & \vdots \\ B_0(s_N) & \cdots & B_K(s_N) \end{bmatrix}$$

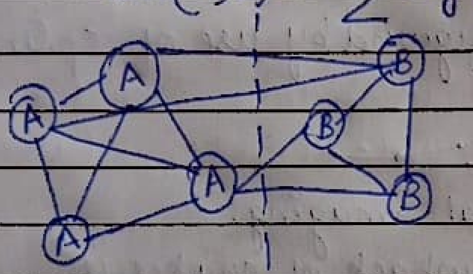
$$X = \begin{bmatrix} x^T(0) \\ \vdots \\ x^T(K) \end{bmatrix}$$

(14) Explain normalised cuts in detail with an example.

→ Normalized cuts technique introduced by Shi and Malik examines the affinities between nearby pixels and tries to seperate groups that are connected by weak affinities.

→ The cut between two groups A and B is defined as

$$cut(A,B) = \sum w_{ij}$$



$$\text{Normalised cut} = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)}$$

$$assoc(A,V) = assoc(A,A) + cut(A,B)$$

→ cuts and associations can be thought of as area sums in the weight matrix W, where the entries of the matrix have been arranged so that the nodes in A come first and the node in B come second.

→ These normalised values better reflect the fitness of a particular segmentation, since they look for collection of edges that are weak relative to all of the edges both inside and emanating from particular region.

→ computing the normalised cut is NP-complete.

(15) Explain the following system

a) scissors

b) level sets

(2) Approaches used to locate boundary curves in images. Explain intelligent Scissors and level set in detail

→ we have three related approaches to locating such boundary curves in image

→ The first, originally called as snakes - is energy-minimizing 2D Spline curve that move towards strong edges.

→ Second one is intelligent scissors allow to sketch a real time curve

→ level set techniques evolve the zero set of characteristic function.

Intelligent Scissors:

→ Intelligent scissors allow objects to be extracted quickly & accurately using simple gesture motions with a mouse.

→ It uses Dijikstra algorithm and Breadth First search algorithm.

→ Navigate the boundary

↓

Rough curve

↓

Freeze

↓

cut.

Level Set Technique:

→ This technique acts as a tool for numerical analysis of surfaces and shapes.

→ For a 3D Space the constant value c in the range is calculated by

$$C = f(x, y, z)$$

→ It is said to have flexible material domain

→ For local measurements we use local minima, in level set.

→ Fastest method.

(16) Categorise the various technique developed for solving pose estimation?

Linear algorithms

The simplest way to recover pose of the camera is to form a set of linear equations analogous to those used for 2D motion

$$x_i = \frac{P_{00} x_i + P_{01} y_i + P_{02} Z_i + P_{03}}{P_{20} x_i + P_{21} y_i + P_{22} Z_i + P_{23}}$$

$$y_i = \frac{P_{10} x_i + P_{11} y_i + P_{12} Z_i + P_{13}}{P_{20} x_i + P_{21} y_i + P_{22} Z_i + P_{23}}.$$

→ The Snake algorithm in computer vision is also called as Active contours.

→ It works like stretched elastic band being released.

→ It is to identify and outline the target objects.

→ Snakes are configured by use of Spline.

→ Used in medical imaging.

Disadv:

→ Noise sensitivity

③ note on pose estimation

→ pose estimation is a c.v. technique to track the movements of the person or objects.

→ These are usually found out by key points.

→ The connection between 2 points is called Pair

→ These are widely used in Human-computer interaction, motion analysis, Sports and fitness and robotics.

Skeleton based model.

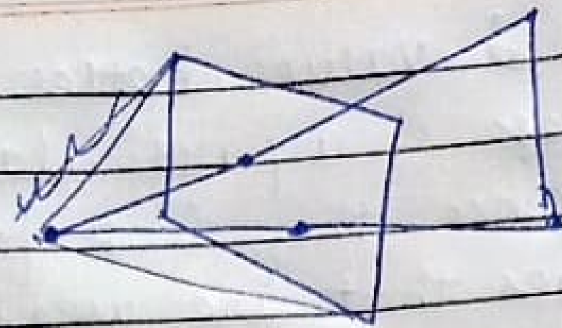
④ Edge linking.

→ Useful technique where the boundaries between objects are automatically identified

→ very useful for segmentation

→ Edge linking

- Gradient based
  - → Sobel operator
- Guassian-based
  - → Canny [illegible]



⊙ Iterative algorithms

The most accurate way to estimate pose is to directly minimise the squared reprojection error for the 2D points as a function of the unknown pose parameters.

$$= \sum \left(\frac{\delta f}{\delta R} \Delta R + \frac{\delta f}{\delta t} \Delta t + \frac{\delta f}{\delta K} \Delta K - r_i \right)$$

(18) Give the equation for the following and briefly summarize each in few words.

a) True +ve rate & false +ve rate (4)

b) +ve predictive value (PPV) & accuracy (6)

Set-B

(6) How can we quantify the performance of a matching algorithm?

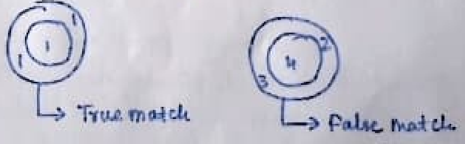
→ Matching algorithm deals with how much the image matches with the other one in order to further process the image to its next step.

→ As a first step we use Euclidean feature descriptor to match the potential matches

→ If the images are not same they have changed too much.

→ Then as next step it tries to match the known matching objects.

→ The transformation to new scaled basis is called whitening.



SRM INSTITUTE OF SCIENCE & TECHNOLOGY

Performance of matching algorithm

TP: True positive

FN: False -ve

FP: False +ve

TN: True -ve.

TPR = True positive rate

$$= \cancel{\frac{TP}{TP+FN}} = \frac{TP}{P}$$

$$FPR = \frac{FP}{N}$$

$$PPV = \frac{TP}{P'}$$

$$ACC = \frac{TP+TN}{P+N}$$