Computer Vision

1)Calculate the gradient of the following function f(x, y) = (x-2)3 sin y at [0,0]

2) In the human retina, the color-sensitive light receptors are called cones.

TRUE FALSE

3) A single rod is sensitive to a single photon of light.

TRUE FALSE

4) The focal length of a camera is the distance between its lens and the object to be photographed

TRUE FALSE

5) Rotation by 30o about x axis followed by 60o about z axis is the same as rotation by 60o about z axis followed by 30o about x axis

TRUE FALSE

6) Mean-Shift algorithm is a filtering technique to smooth an image

TRUE FALSE

7) A Gaussian filter performs very well in cases of salt and pepper noise.

TRUE FALSE

8) The rank of fundamental matrix is 3

TRUE FALSE

9) With same initialization, k-means always converge to same solution

TRUE FALSE

10)Convolution is equivalent to cross-correlation without flips.

TRUE FALSE

11) SIFT descriptions is not invariant because it is rotated based on the keypoint orientation.

TRUE FALSE

→ Many 3D points in world can be mapped to the same 2D point in image TRUE

→ Blue is the best light component to penetrate underwater TRUE

→ Gaussian filter performs very well in the cases with salt and pepper noise FALSE

→ Average (box) filter performs very well in the cases with salt and pepper noise FALSE

→ Rank of fundamental matrix is 3 FALSE

→ With same initialization, k-means always converge to same solution

→ In Hough transform a line in x, y space is transformed into point in a, b space FALSE

A single point in x, y space gives a line in (a, b) space.

→ Correlation is a measure of relatedness of two signals TRUE

→ In a perspective projection of a scene lines that are parallel in the world always project to lines that are parallel in image FALSE

* How many degrees of freedom are there in 2D rotation matrix?

→2 →1 →3 →4

* How many degrees of freedom are there in 3D rotation matrix?

→4 →3 →2 →1

* If the essential matrix for point p1 and p2 is E. Essential matrix for p2 and p1 is:

→ET →E-1 →E →it’s different with no relation to E

* Degrees of freedom of essential matrix, why

→ 2 : the rank of the essential matrix is 2

→ 7 : 3 dof for rotation, 3 dof for translation, up to a scale, so 1 dof is added

→ 6 : 3 dof for rotation, 3 dof for translation

→ 5 : 3 dof for rotation, 3 dof for translation, up to a scale, so 1 dof is removed

* If a point P has value X, Y, Z in 3D space in camera coordinates , and the camera is in standard position (center down z axis) which of the equations relates to the projection of this point P into camera coordinates x and y?

→ x = f Y/Z y = f X/Z

→ x = f Z/X y = f Z/Y

→ x = XZ/f y = YZ/f

→ x = fX/Z y = fY/Z

17) the image A is the original image. To obtain image B, we added some white noise to A. three vision engineers performed canny edge detection on image B( Let’s assume we can disregard the hysteresis step), as aprt of their job interview process. Images C,D,E are their final edge detection results. Only applicant correctly answered the question. Explain for the other two images what needs to be done in order to get good edge detection result.

A- e rregullt (nje goce)

B- me ca noise, me ca miza

C- e zeze , thin vija te bardha

D- komplet moza, asgje nuk duket, vtm ca viza te bardha(silueta e goces)

E- e zeze, vija te bardha te trasha

* A → original image

B → added some white noise to A (ne foto pak si me e kriposur)

Perform Canny edge detector. Imazhi i sakte (zgjidh figuren dhe shpjego)

→ C (duket si te detyra)

→ D (nuk duket asgje, si gri me miza)

→ E (me vija shume te trasha ne krahasim me C)

1. Consider the following 8x8 picture:

1 1 3 2 2 0 3 2

0 3 0 1 3 1 1 1

2 3 10 7 9 1 2 1

0 2 2 7 8 9 0 0

2 1 7 9 8 8 0 2

3 2 7 9 7 7 2 2

1 3 2 1 3 0 0 0

1 1 1 3 2 0 0 0

1. Which of these two mages on the top represents the correct image histogram?
2. What could be a suitable threshold that can be used to obtain a good binary image replica corresponding to the grey scale image provided above (more than 1 answer may be correct)

10

1.00 2.00 3.00 4.00 5.00 6.00 7.00 8.00 9.00 10 10

12

10

8

6

4

2

0

1 3 5 7 9 11 13 ………………………………………………………. 63

Find correct image histogram (2 zgjedhje me figura)

→ Njeri eshte i trashe me vlera nga 0 ne 10

→ Tjetri eshte i holle me vlera nga 1 ne 63

Find suitable threshold to have good binary image replica corresponding to gray scale image provided above

18) Write a 3 by 3 pixel linear filter that shifts an image 1 pixel upwards and also reduces the image brightness by 50%

\_\_ \_\_ \_\_

\_\_ \_\_ \_\_

\_\_ \_\_ \_\_

20)Use single link agglomerative clustering to group the data described by the following distance matrix. Which of the following clustering is correct?

| A | B | C | D

-------|------|------|------|------------

A | 0 | 1 | 4 | 5

B | | 0 | 2 | 6

C | | | 0 | 3

D | | | | 0

A B C D A B C D

A B C D A B C D

21) Use complete link agglomerative clustering to group the data described by the following distance matrix. Which of the following clustering is correct?

(si ajo lart)

22)The organisers of the Bremen market Christmas have hired you as a computer vision Engineer, to support them in various operations. Only people with a visible entry bracelet can access the area and the organisers want this check to be performed without human intervention, which would otherwise create a barrier and a queue. In order to minimize contact, when a person wants to buy smth, the market organizers want a system which would recognize the person and automatically deduct the amount from their bank account. Finally, in a case a person is later diagnosed with covid19, they want to be able to contact all the persons who have been near in order to inform them about the risks ( note that contact tracing apps cannot be used, as there is no requirement of having a functioning mobile phone to walk in the Market.) What are he steps you would do. Please clarify all assumptions. Focus on comp.Vison aspects only.(eg. No financial transaction implementation.)

Requirements:

Allow entering for people with visible bracelet

Recognize person automatically and deduct amount from bank account

Keep track of people near each single person to contact later if this single person has covid

CV

1. A single rod is sensitive to a single photon of light.

T F

2. The focal length of a camera is the distance between its lens and the object to be photographed.

the distance measured in millimeters, between the optical center of the lens and the camera sensor, where the light information is recorded

T F

3. Is matrix multiplication commutative?

T F

4. Pose is an intrinsic property of an object.

extrinsic

T F

5.Rotating by 30 deg about x-axis and then 60 deg about z-axis is the same as rotating by 60 deg about z-axis followed by 30 deg about x-axis?

T F

6. Pose is an intrinsic property of an object.

T F

7. Blue is the best light component to penetrate underwater.

T F

8. The rank of the fundamental matrix is 3

T F

9. Two images are taken by a single perspective projection camera that has been rotated about the optical center. The scene depth can therefore be recovered.

T F

10. K-means can get stuck at local minima.

T F

11. If a point P has value X, Y, Z in 3D space in camera coordinates, and the camera is in the standard position (center down the z-axis), which of the following equations relates the projection of this point P into camera co-ordinates x and y?

A. x=f Z/X

y= f Z/Y

B. x=f X/Z

y=f Y/Z

C. x=f Y/Z

y= f X/Z

D x=X Z/F

y= Y Z/f

12. How many degrees of freedom are there In a 2D rotation matrix

1 2 3 4

13. How many degrees of freedom are there In a 3D rotation matrix

1 2 3 4

14. If the essential matrix for point pair p1 and p2 is E, what is the essential matrix for point pair p2 and p1?

E^T It is a different matrix with no relation to E E^(-1) E

15. What are the degrees of freedom of the essential matrix and why?

2: the rank of the essential matrix is 2

6: 3 dof for. Rotation 3, for translation

7: 3 dof for. Rotation 3, for translation. Up to scale so 1 is added

5: 3 dof for. Rotation 3, for translation. Up to scale so 1 is removed

16. The image A is the original image. To obtain image B, we added some white noise to A. The canny edge detection on image B(assume we can disregard the hysteresis step), are images C, D, E are their final edge detection results. Only one applicant correctly answered. Which is the correct result?

Explain for the other two images what needs to be done to get a good edge detection

The optimal edge detector should minimize the probability of false positives (detecting suspicious edges caused by noise), as well as that of false negatives (missing real edges) D

The edges must be as close as possible to the true edges

The detector must return one point only for true edge points ; that is, minimize the number of local maxima around the true edge. Picture E

The correct result is image C. the other two d, and d have not successfully applied the cany edge detector or have missed some steps in the process. Multiple algorithms can be used to detected the edges (with the information d and e have) in the images that have failed so, but I will continue with the canny edge detector as it is good and efficient.

For d: seems the student has forgotten to remove the noise, so the first thing to be done is to remove the noise which we can do my convolving the image with a mask that represents a low-pass filter or smoothing operation.

Supposing the result that we got is image b, than we should suppress the image to get rid of the noise that was added from a, this can again be done with the x-y derivatives of gaussian

Then we should compute the orientation and magnitude for the pxels

We should than apply non-max suppression to thin the wide multipixel down to single pixel width.

Using hysteresis and connectivity analysis we can detect edges.

For e: seems the person hass successfully done the nose suppression and computing the gradient magnitude and direction, but seems he has been stuck in suppressing the wide edges,

So he should apply non-maximum suppression – assuming minimal response

Then apply hysteresis and connectivity analysis to detect edges.

17. Write 3x3 matrix filter that shifts the image 1 pixel upwards and reduces the image brightness by 50%

18.

1 1 3 2 2 0 3 2

0 3 0 1 3 1 1 1

2 3 10 7 9 1 2 1

0 2 2 7 8 9 0 0

2 1 7 9 8 8 0 2

3 2 7 9 7 7 2 2

1 3 2 1 3 0 0 0

1. 1 1 3 2 0 0 0
2. Which of these two mages on the top represents the correct image histogram?
3. What could be a suitable threshold that can be used to obtain a good binary image replica corresponding to the grey scale image provided above (more than 1 answer may be correct)

19. Use single link agglomerative clustering to group the data described by the following distance matrix. Which of the following clustering is correct?

| A | B | C | D

-------|------|------|------|------------

A | 0 | 1 | 4 | 5

B | | 0 | 2 | 6

C | | | 0 | 3

D | | | | 0

A B C D A B C D

A B C D A B C D

20. Use complete link agglomerative clustering to group the data described by the following distance matrix. Which of the following clustering is correct?

| A | B | C | D

-------|------|------|------|------------

A | 0 | 1 | 4 | 5

B | | 0 | 2 | 6

C | | | 0 | 3

D | | | | 0

A B C D A B C D

A B C D A B C D

21. Calculate the gradient of the following

(x,y)=(x-2)^3siny at (0,0)

22. Calculate the gradient of the following

(x,y)=e^x (2y^2 +1)^3 at (0,0)

23. Due to the coronavirus pandemic, the city of Bremen has introduced new restrictions for accessing Burgerpark.

Groups of people are allowed up to a maximum of n people, n is 4

The distance between each person in the same group needs to be at least d1 m, d1 now is 1.5 m

-The distance between each group needs to be at least d2 meters, d2 is currently 5 m

In order to enforce these rules, they ask you to fly a drone in the park, which would monitor the composition and trigger a warning when the rules are not respected. Assume that you already have a module to recognize people. What are the steps you would do? Please clarify all assumptions.

Requirement:

Max n in a group 4

D1 inside the group 1.5m

D2 between groups 5m

Fly a drone

Track movements

Trigger warning

Module to recognize people is implemented.

Given the requirements of the “system” to be designed. First I event based and rgb cameras will be needed as we need to track movements in milliseconds and also recognize color with simple cameras.

Because the logic is to give warning, the best scenario would be that the structure or plan is ready and who deviates triggeres a warning, so even though the agglomerative clustering could be used, than we would have to check for the instances (people) not in a cluster as they would be the ones triggereing the alert, but this would create a lot of overhead with 2 thresholds and the cost would be high assuming the park is crowded.

So, fist I would identify each person with the module, and check if there is someone else in their circle of radius 1.5 m.

Super crowded scenario.

Than I would divide the park in zones with a 5m distance from each other .

In each of these zones count the number of people and measure that the distance between them is min 1.5m. If n is greater than 4 or the distance is less than 1.5m trigger a warning for that group, maybe by adding light or sound to them.

For inter-cluster waring, using spatial hashing we can reduce the number of calculations by dividing the park into grids and calculate the distance for grups in neighhbouring grids.

Not so crowded scenario.

K-means would be efficient, but it assumes fixed number of clusters, which would make dynamic implementation difficult.

We can use DBSCAN, which automatically determines the number of clusters based on a threshold. This algo. Also identifies noise points and handles dynamic clustering better. Can improve our system in cases of splitting or merging clusters.

Then cluster for the threshold of distance 1.5 between the elements of the clusters.

Assuming those groups will continue to stay together in the park, we’ll keep track of their movements with the event-based cameras, and if someone shows in the 1.5m radius of the circle for each element of the cluster, or another cluster comes closer than 5m, than trigger a warning. For this we use event-driven processing. To effectively track everyone, use multi-object tracking.

Adding to rgb cameras use stereo cameras for precise 3d positioning and information.

Use **edge computing** to process data locally on drones and only send critical updates to a central system.

Of corse there should be longer analysis of the relation and efficiency of this 1.5m in-cluster distance and 5m inter-cluster distance to make sure what the warning should be made for.