NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

OLLSCOIL NA hÉIREANN MÁ NUAD

THE NATIONAL UNIVERSITY OF IRELAND MAYNOOTH

AUTUMN 2014 EXAMINATION

CS410

Computer Vision

Dr. D. Charles, Dr. A. Winstanley, Mr. J. McDonald

Time allowed: 2 hours

Answer ***three*** questions

All questions carry equal marks

[25 marks]

1 (a) The 3D to 2D projection performed by a camera can be modelled as a linear system as follows: [10 marks]

$$\mathbf{p} = \mathbf{K}(\mathbf{R} \quad \mathbf{t})\mathbf{P}$$

Here, $\mathbf{P}$ is a homogeneous 3D world point, and, $\mathbf{p}$ is the corresponding projection of that point in image space. Provide an explanation of the properties of the camera that are modelled by matrix $\mathbf{K}$, the matrix $\mathbf{R}$ and the vector $\mathbf{t}$.

The expression below shows the structure of $\mathbf{K}$:

$$\mathbf{K} = \begin{pmatrix} kf & 0 & u \\ 0 & lf & v \\ 0 & 0 & 1 \end{pmatrix}$$

What aspect(s) of the transformation from 3D world space to 2D image space are modelled by each of the elements of $\mathbf{K}$? In your answer you should explain the meaning of each of the elements of the matrix.

(b) An alternative view considers the camera projection as a single 3×4 matrix (i.e. it does not decompose the transformation into $\mathbf{K}(\mathbf{R} \quad \mathbf{t})$). Here the projection of the camera system is modelled directly as: [10 marks]

$$\mathbf{p} = \mathbf{MP} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix}\mathbf{P} = \begin{pmatrix} \mathbf{m}_1^T \\ \mathbf{m}_2^T \\ \mathbf{m}_3^T \end{pmatrix}\mathbf{P} = \begin{pmatrix} \mathbf{m}_1^T\mathbf{P} \\ \mathbf{m}_2^T\mathbf{P} \\ \mathbf{m}_3^T\mathbf{P} \end{pmatrix}$$

In the course we saw how it is possible to calibrate for $\mathbf{M}$ by first collecting a set of 3D world point to 2D image point correspondences. Show how these constraints may be used to construct a homogeneous linear system of the form:

$$\mathbf{Ax} = \mathbf{0}$$

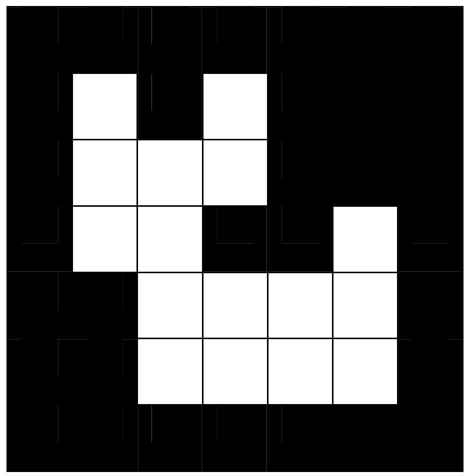
Given this linear system provide a detailed explanation of the set of steps involved in estimating the camera matrix.

(c) Given that the 3D world point, (10.0,20.0,10.0), corresponds to the 2D image point, (50.0,100.0), compute the first **two** rows of the matrix $\mathbf{A}$ above. [5 marks]

[25 marks]

2 (a) Give details of the sequential 2D labelling algorithm for labelling the individual components in a binary image. You should provide a pseudo-code description of the algorithm in your answer. [9 marks]

With the aid of diagrams explain how the algorithm would label following binary. Note your diagrams should show the result of each stage of processing, including both the intermediate image representation after the first labelled image. You should also show the state of the equivalence table after phase 1.

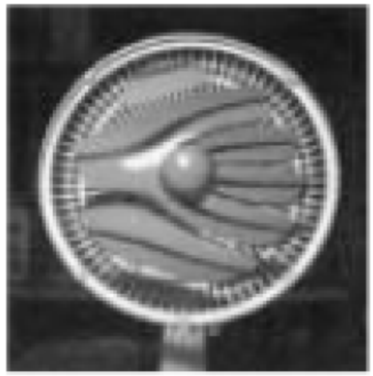


(b) Provide expressions for the zeroth and first moments of the foreground pixels in a binary image. [8 marks]

Using these expressions calculate the zeroth and first moment of the object in the image shown in part (a).

Question 2 continued overleaf…

(c) The figure below shows an image of a fan spinning at high-speed. As can be seen from the figure, the image is significantly distorted. What type of camera sensor was used to capture this image and what is the term given to this type of distortion? Provide an explanation for the source of this type of distortion. Finally list and explain two mechanisms for overcoming this type of distortion. [8 marks]



[25 marks]

3 (a) Explain the template matching paradigm to detecting known objects in images. [7 marks]

(b) Central to the application of the template matching paradigm is the use of a similarity metric. Shown below are expressions for the four principal similarity metrics covered in the course. [12 marks]

$$E(m,n)=\sqrt{\sum_i \sum_j [g(i,j)-t(i-m,j-n)]^2}$$

$$S(m,n)=\sum_i \sum_j |g(i,j)-t(i-m,j-n)|$$

$$R(m,n)=\sum_i \sum_j g(i,j)t(i-m,j-n)$$

$$N(m,n)=\frac{R(m,n)}{\left(\sqrt{\sum_i \sum_j g(i,j)^2}\sqrt{\sum_i \sum_j t(i-m,j-n)^2}\right)}$$

For each metric provided identify its name, and provide an intuitive explanation of the how it characterises similarity. Finally compare each of the metrics in terms of their effectiveness.

(c) Given the three sub-windows shown below, what would $N(m,n)$ evaluate to if it were used to compare sub-windows (a) and (b)? What would it evaluate to if it were used to compare (a) and (c)? Note you should justify your answers either by direct calculation or by providing an explanation for your answer. [6 marks]

100	0	100
0	100	0
100	0	100

(a)

50	0	50
0	50	0
50	0	50

(b)

0	100	0
100	0	100
0	100	0

(c)

[25 marks]

4 (a) What is the scale-space of an image and what is its importance in SIFT's approach to interest point detection? Given a particular SIFT feature point, how does the SIFT algorithm compute its canonical orientation? [10 marks]

(b) Consider a scenario where you are given three images of individual book covers. You are also given a number of images of cluttered scenes, some of which contain instances of a subset of the books. [15 marks]

Give details of an algorithm for automatically indentifying which images contain which books.

Suggest an approach to locating and segmenting the books in the image.