—— SOLUTIONS ——

King's College London

Degree Programmes BSc, MSci

Module Code 6CCS3COV

Module Title Computer Vision

Examination Period January 2016 (Period 1)

Time Allowed Three hours

Rubric ANSWER QUESTION ONE AND ANY THREE OTHER QUESTIONS.

All questions carry equal marks. If more than three questions other than Question One are answered, clearly indicate which answers you would like to be marked. Write this clearly in the dedicated section on the front page of the answer booklet. If you do not indicate which questions you would like to be marked, Question One and the first three subsequent questions answered, in exam paper order, will be marked.

Calculators Calculators may be used. The following models are permitted: Casio fx83 / Casio fx85.

Notes Books, notes or other written material may not be brought into this examination

PLEASE DO NOT REMOVE THIS PAPER FROM THE EXAMINATION ROOM

1. Compulsory question

a. What is the correspondence problem? Describe one way in which the correspondence problem is similar for stereo vision and video streams, and one way in which it is different.

[6 marks]

Answer

The correspondence problem is the problem of finding matching image elements across different views of the same scene.

In both stereo and video, we are looking for the same objects in multiple frames, so we are solving the same problem in both cases.

In stereo we are looking at exactly the same view, but from a (slightly) different position. In video we are looking at a stream of views of the same scene, albeit one which may have changed slightly from one frame to another.

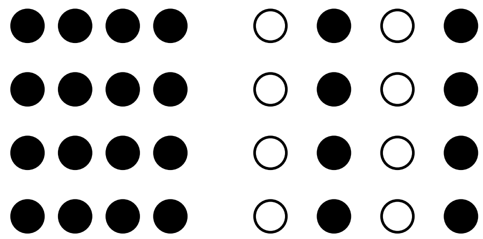
Marking scheme

2 marks for the definition. 2 for the similarity and 2 for the difference.

This question is on computational mid-level vision

b. For the two following images, describe the grouping that you observe, and the Gestalt Law that leads to that grouping.

[4 marks]



(a) (b)

Answer

In the left figure, we see the dots grouped as four horizontal rows. This is the law of *proximity*.

In the right figure we see columns of black and white. This is the law of *similarity*.

Marking scheme

2 marks for each law that correctly relates to the grouping described.

This question is on biological mid-level vision

c. Give an example of the way in which computer vision is an "ill-posed" problem. What is the practical consequence of this issue?

[5 marks]

Answer

Computer vision is "ill posed" because there are typically many different scenes that could give rise to a given image, so interpreting the image involves picking one scene from many with insufficient information.

One practical consequence is that intrepreting an image always requires the use of some prior information about the contents of the image.

Marking scheme

3 marks for describing "ill-posed", two for the consequence.

This question is on computer vision as a whole.

d. What is the result of the convolution of mask H with image I?

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad I = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 1 & 0 & 0.5 & 0 \\ 0 & 1 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

The result should be an image that is the same size as I.

[5 marks]

Answer

$$\begin{bmatrix} 1 & 0.5 & 1 & 0 & 0 \\ 1 & 2 & 0.5 & 1.5 & 0 \\ 0 & 2 & 3 & 0 & 1 \\ 0 & 0 & 2 & 2 & 0 \end{bmatrix}$$

Marking scheme

One mark for each correct column.

This is a question on computational low-level vision.

e. What is the Lateral Geniculate Nucleus, and what role does it play in the human visual system?

[5 marks]

Answer

Lateral Geniculate Nucleus (LGN) is a point along the visual pathway between the retina and the primary visual cortex.

The optic nerves start from the retina and end in the LGN and Superior Colliculus after crossing at the optic chiasma. The axons coming from the ganglion cells in the eye form synapses on neurons in the LGN and Superior Colliculus. The axons of the LGN neurons project into the primary visual cortex.

The optic nerve from the left eye contains information about both the lefthand and righthand fields of view, and the optic nerve from the right eye contains information about both the lefthand and righthand field of view. In the LGN the relevant information is "sorted" so that all the information from the lefthand field of view is carried to the same area of the primary visual cortex and similarly for those from the righthand field of view.

Marking scheme

5 marks for the description above or its equivalent.

This is a question on biological low-level vision.

2.

a. What is a "centre-surround receptive field" and what does it detect?

[5 marks]

Answer

A receptive field is an area of visual space, for example an area on the retina, from which a neuron receives input.

A centre-surround receptive field is one that detects a difference between the centre of the area and the outside of the area. Thus a neuron with a centre-surround receptive field will respond either if the centre is brighter than the surround (an on-centre, off-surround cell) or if the surround is brighter than the centre (and off-centre, on-surround cell).

Marking scheme

Proportionally, depending on how much of the answer above is given.

This is a question on biological low-level vision.

b. Explain how ganglion cells in the eye and simple cells in the primary visual cortex (V1) work together to detect edges.

[10 marks]

Answer

Ganglion cells are centre-surround, so detect (for example) darker areas surrounded by lighter areas. The retina is over-represented, so that every area in the retina is part of the receptive fields of several ganglion cells. Where several off-centre, on-surround receptive fields are lined up, their respective ganglion cells will all detect something when a dark line on a lighter background lies on their combined receptive field — where there is an edge.

Simple cells in V1 are connected in just this way — a single simple cell receives input that orignates in the ganglion cells of receptive fields that overlap along a line in a specific orientation. When an edge lines up with the aligned receptive fields, the simple cell fires and the edge is detected. This describes the detection of dark lines on light backgrounds. If the initial receptive fields are on-centre, off-surround, the simple cell will detect light lines on dark backgrounds.

Marking scheme

4 marks for describing what ganglion cells detect, 4 marks for describing how simple cells combine adjacent receptive fields to create edge detectors, and 2 marks for extending the description to deal with the two kinds of receptive field and their relation with light and dark detection.

This is a question on biological low-level vision.

c. Give an example of the function of complex cells in the primary visual cortex (V1).

[5 marks]

Answer

Complex cells combine input from several simple cells, so can detect patterns that are combinations of single cell detections. For example, complex cells can detect edges in a (somewhat) position independent way, by taking input from simple cells that detect an edge at different points in the visual field.

Marking scheme

3 marks for the way that complex cells work, two for an example of what the cells can detect.

This is a question on biological low-level vision.

d. Describe how lateral connections in the primary visual cortex (V1) can explain some of the Gestalt laws.

[5 marks]

Answer

Lateral connections in V1 connect together areas of V1 that deal with adjacent areas of the visual field. If these conenctions are between two areas of V1 that are stimulated by the same visual cue, for example a line segment, then the effect of the connection is to make two adjacent line segments stand out more than either on its own. When the line segments are side-by-side, this explains the law of similarity. When the line segments are end-on, this explains the law of continuity.

Marking scheme

Two marks for explaining what lateral connections do, two for the first example of the connection to a Gestalt law, one for any further examples.

This is a question on biological mid-level vision.

3.

a. One approach to image recognition is template matching. Describe briefly in words, or using pseudocode, each step required to perform template matching on an image.

[5 marks]

Answer

First, create a template, an image of the object that is to be recognized
For every template:
    Search every image region
    Calculate similarity between template and image region
    Choose the "best" match, or all matches where similarity exceeds a threshold

Marking scheme

1 mark per line.

This is a question on computational high-level vision.

b. Here is a template T and an image I.

$$T = \begin{bmatrix} 250 & 200 & 250 \\ 200 & 50 & 150 \\ 200 & 150 & 100 \end{bmatrix} \qquad I = \begin{bmatrix} 50 & 150 & 150 & 200 \\ 100 & 200 & 50 & 50 \\ 50 & 100 & 100 & 150 \\ 50 & 50 & 50 & 100 \end{bmatrix}$$

Calculate the result of performing template matching on the image using sum of absolute differences (SAD) as the similarity measure.

Assuming that the image contains exactly one instance of the object in the template, suggest the location of the centre of the object in the image.

[10 marks]

Answer

There are four locations where the template can be matched to the object — with the centre of the template (the 50) over location (2, 2), in MATLAB terms (over the top left 200), (2, 3), (3, 2) and (3, 3).

Here are the similarities, for (2, 2):

$$\begin{aligned}
SAD_{2,2} &= |250-50| + |200-150| + |250-150| \\
&\quad + |200-100| + |50-200| + |150-50| \\
&\quad + |200-50| + |150-100| + |100-100| \\
&= 200 + 50 + 50 \\
&\quad + 100 + 150 + 100 \\
&\quad + 150 + 50 + 0 \\
&= 850
\end{aligned}$$

Similarly,

$$\begin{aligned}
SAD_{2,3} &= 400 \\
SAD_{3,2} &= 900 \\
SAD_{3,3} &= 800
\end{aligned}$$

Since we assume that there is exactly one match, the match is at (2, 3).

Marking scheme

2 marks for the correct value of each template position, and 2 for the location of the match.

This is a question on computational high-level vision.

c. A computer vision system uses template matching to perform object recognition.

i. The system needs to detect 20 different objects each of which can be seen from 10 different viewpoints. How many templates will be required? Why?

[4 marks]

Answer

Each viewpoint will typically require a separate template, so 200 templates.

Marking scheme

2 marks for 200, 2 for the reason.

This is a question on computational high-level vision.

ii. If each template is 9 pixels by 9 pixels, an image is 300 pixels by 200 pixels, and comparisons are done using sum of absolute differences, how many floating point operations are required to process one image?

[6 marks]

Answer

The first position for a 9 by 9 template is with the centre at (5, 5), and the last position on the first line is 296, 5), so there are 292 comparisons on each line. The last line will be the 195th, so there are 192 lines of comparison.

Thus there are $292 \times 192 = 56064$ comparisons.

At each location there is a subtraction (so 81 operations), and an absolute (another 81 operations) and then these results are summed (requiring 81-1 operations). So there are 81+81+80=242 operations per comparison.

Thus we have 56064x242=13,567,488 floating point operations to compare one template with the image. Comparing all the templates will require: 200x13,567,488=2,713,497,600 floating point operations.

Marking scheme

3 marks for the calculation of comparisons, 3 for the calculation of operations, approximately correct calculations are acceptable.

This is a question on computational high-level vision.

4.

a. Explain the effect of convolving an image with the box mask:

$$\begin{bmatrix} 0.04 & 0.04 & 0.04 & 0.04 & 0.04 \\ 0.04 & 0.04 & 0.04 & 0.04 & 0.04 \\ 0.04 & 0.04 & 0.04 & 0.04 & 0.04 \\ 0.04 & 0.04 & 0.04 & 0.04 & 0.04 \\ 0.04 & 0.04 & 0.04 & 0.04 & 0.04 \end{bmatrix}$$

[4 marks]

Answer

The mask will have an averaging effect, replacing the value of a pixel with the average of the values of surrounding pixels covered by the mask. This will blur the image, reducing noise. Since the mask elements add to 1, the mask will not change the intensity of the image.

Marking scheme

2 marks for remarks around "averaging" or blurring, and 2 marks for touching on normalisation.

This is a question on computational low-level vision.

b. Write down a difference mask, and explain the effect of convolving an image with that mask. Contrast your answer with the your answer to Question 4.a.

[6 marks]

Answer

A difference mask is $[1, -1]$, and it computes the difference between two pixels in the x direction. Convolving an image with a difference mask will highlight changes in intensity in the direction that the mask works (in the x direction in this case). These typically occur at edges of objects. Whereas a box masks averages, reducing the difference between adjacent pixels, a difference mask detects the difference between two pixels.

Marking scheme

2 marks for knowing what a difference mask does. 2 marks for the result of convolution with the mask. 2 marks for the contrast between box and difference masks.

This is a question on computational low-level vision.

c. A Laplacian mask can be thought of as a combination of several difference masks.

i. Write down a 3 by 3 Laplacian mask and describe the effect of convolving an image with a Laplacian mask?

[6 marks]

Answer

A Laplacian mask is:

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

and broadly it is a difference mask in every direction — x and y and diagonal. The effect of convolving an image with a Laplacian mask is to detect discontinuities in intensity in every direction — x direction, y direction and diagonals — identifying all the edges in an image.

Marking scheme

2 marks for the mask. 4 for describing the effect of the convolution.

This is a question on computational low-level vision.

ii. Write down one advantage and one disadvantage of using a Laplacian mask to process an image.

[4 marks]

Answer

The Lapalacian is very good at picking up discontinuities. However, it is most senstive to a single pixel that stands out from its surrounding, so every noise pixel will be indentified, leading to many false positives when doing edge detection.

Marking scheme

2 marks for the advantage, 2 for the disadvantage.

This is a question on computational low-level vision.

iii. How can the disadvantage you gave in your answer to Question 4.c.ii be mitigated?

[5 marks]

Answer

The main disadvantage to using the Laplacian is the fact that it identifies noise as a discontinuity, so the way to mitigate this is to apply an averaging filter before applying the Laplacian. This is typically done by convolving together a Laplacian mask and a Gaussian mask to obtain the "Laplacian of Gaussian" or LoG mask.

Marking scheme

3 marks for the use of averaging before applying the Laplacian. 2 marks for discussing the Laplacian of Gaussian.

This is a question on computational low-level vision.

5.

a. In mid-level vision, what do the terms "grouping" and "segmentation" mean?

[4 marks]

Answer

Grouping is grouping together those elements of an image that belong together. Segmentation is distinguishing those elements that belong together in an image from other elements.

Marking scheme

2 marks for each definition.

This is a question on computational mid-level vision.

b. One approach to segmentation is region splitting and merging. Describe briefly in words, or using pseudocode, each step required to perform segmentation by region splitting and merging.

[7 marks]

Answer

Start with the whole image labelled as a single region
For each region:
    If all pixels are not similar, split the four quadrants into different regions.
    Continue until each region is homogeneous.
For each region:
    Compare to neighbours and merge neighbouring regions which are similar.
    Continue until no more regions can merge.

Marking scheme

1 mark for each line in the algorithm.

This is a question on computational mid-level vision.

c. The array below gives the feature vectors for each pixel in a 4×4 image:

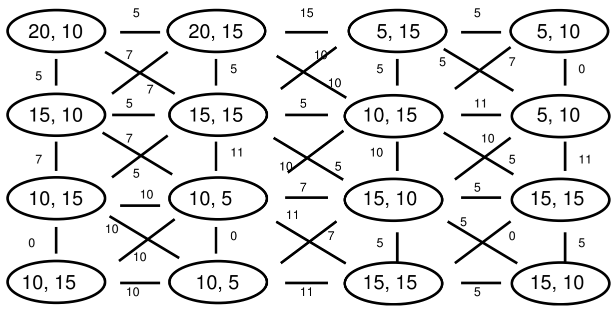
$$\begin{bmatrix} (20,10) & (20,15) & (5,15) & (5,10) \\ (15,10) & (15,15) & (10,15) & (5,10) \\ (10,15) & (10,5) & (15,10) & (15,15) \\ (10,15) & (10,5) & (15,15) & (15,10) \end{bmatrix}$$

Use region splitting and merging to assign each pixel to a region under the assumptions that (1) Euclidian distance is used as the method to assess similarity between regions; (2) the threshold for deciding that regions are similar is 9; (3) regions have horizontal, vertical and diagonal neighbours.

[12 marks]

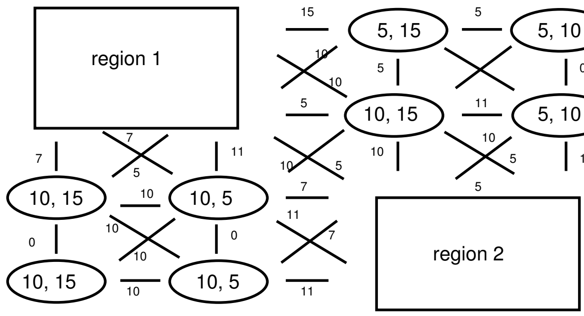
<u>Answer</u>

We start by labelling the whole image as one region and testing whether the similarity between any pair of pixels exceeds the threshold. These are the pairwise distances computed by Euclidian distance:



Clearly some of these distances exceed 9 so we split into quadrants.

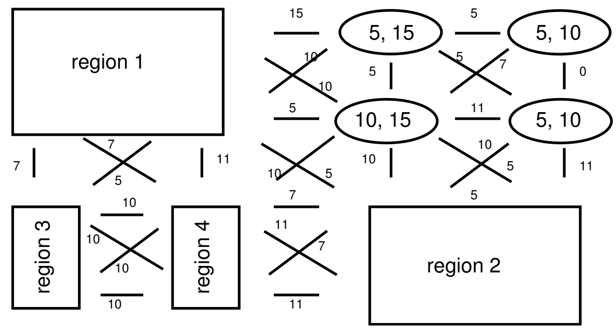
The top left quadrant is now all below the similarity threshold, as is the bottom right, so we have an initial set of regions that looks like this:



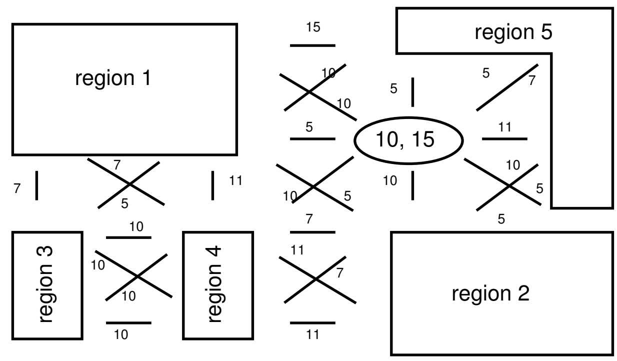
We then split bottom left and top right into quadrants, which this time are individual pixels. This ends the splitting phase.

Next we merge.

We merge into regions all pixels with neighbours that are under the threshold. Thus the bottom left splits into two:

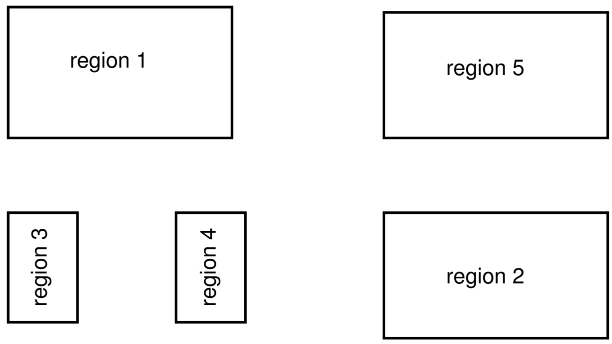


The three top-right pixels that have 5 as the first value all merge to give:



The last pixel on the top right is them compared with the four adjacent regions, computing the distance to the average value for that region.

Though the distance to individual pixels in the regions (shown in the diagram) exceed the threshold, the distance between the last lone pixel and regions 1, 2 and 5 are all below the threshold, so it can be merged with any of these. Here we consider it is merged with region 5 to give:



Then we consider merging regions by comparing the average value across a region with the threshold. Here, regions 1 and 2 will merge, and further rounds of merging may occur.

Marking scheme

4 marks for demonstrating how splitting works, 4 marks for demonstrating how merging works.

This is a question on computational mid-level vision.

d. Give one weakness of region splitting and merging for segmentation.

[2 marks]

Answer

Splitting and merging works well if the regions are fairly homogeneous. If not, many spurious regions are created.

Marking scheme

2 marks for the answer.

This is a question on computational mid-level vision.

6.

a. Explain how a CCD camera forms an RGB image.

[5 marks]

Answer

A CCD camera is an array of MOS capacitors. Light falling on such a capacitor causes it to accumulate charge in proportion to the light intensity, and measuring this produces a picture.

Individual diodes are made sensitive to R, G and B light by placing a filter between the light source and the diode. Cameras can either use three filters and three arrays of diodes — to collect R, G and B values directly for each pixel — or can use a mask that provides R, G or B filters for individual diodes and interpolate between the recorded values to estimate the missing readings.

Marking scheme

Two marks for recalling that a CCD camera is an array of photodiodes that accumulate charge, and three marks for recounting how different colour channels are detected.

This is a question on image formation.

b. Given an RGB image, how would you create a greyscale image from it? Your answer should explain how the computation is carried out, it should not just give the relevant MATLAB command.

[3 marks]

Answer

To compute grayscale from RGB, one can just average the R, G abd B values for each pixel. Weighted averages are also approriate.

Marking scheme

3 marks for the above.

This is a question on image formation.

c. The raw output from a CCD device is the array of values below:

G11	R12	G13	R14	G15
B21	G22	B23	G24	B25
G31	R32	G33	R34	G35
B41	G42	B43	G44	B45
G51	R52	G53	R54	G55

where R12, R14 and so on are red values; G11, G13 and so on are green values; and B21, B23 and so on are blue values.

Describe how to compute red, green and blue values for each array element, and comment on how well you think the method you describe will approximate the input to the CCD.

[5 marks]

Answer

There are a number of possible schemes. For example, we can interpolate the value of a pixel from those surrounding it. This works well for central pixels — the G value for pixel 23 can be the average of the G values for 13, 33, 22 and 24 — but less well for edge pixels, for example, computing an R value for pixel 21.

Marking scheme

3 marks for stating a mechanism for computing values, and 2 for critiquing it.

This is a question on image formation.

d. A point in 3D space has coordinates [20,20,600] mm relative to the camera reference frame. If the image principal point is at coordinates [250,150] pixels, and the magnification factors in the x and y directions are 800 and 600, then determine the location of the point in the image. Assume that the camera does not suffer from skew or any other defect.

[8 marks]

Answer

The key equation here is:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z} \begin{pmatrix} \alpha & 0 & o_x \\ 0 & \beta & o_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

slotting in the values from the question, we get:

$$\begin{aligned}
\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} &= \frac{1}{600} \begin{pmatrix} 800 & 0 & 250 \\ 0 & 600 & 150 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 20 \\ 20 \\ 600 \\ 1 \end{pmatrix} \\
&= \frac{1}{600} \begin{pmatrix} 800 & 0 & 250 \\ 0 & 600 & 150 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 20 \\ 20 \\ 600 \end{pmatrix} \\
&= \frac{1}{600} \begin{pmatrix} 800 \times 20 + 0 + 250 \times 600 \\ 0 + 600 \times 20 + 150 \times 600 \\ 0 + 0 + 1 \times 600 \end{pmatrix} \\
&= \begin{pmatrix} 277 \\ 170 \\ 1 \end{pmatrix}
\end{aligned}$$

Note that the 277 above is rounded from 276.6667. Either answer is fine.

Marking scheme

4 marks for the equation, 2 for knowing how to fill in the values, 2 for the correct answer.

This is a question on image formation.

e. How do the coordinates of the image change if the object moves to the location [20,20,900]?

[4 marks]

Answer

$$
\begin{aligned}
\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} &= \frac{1}{900}\begin{pmatrix} 800 & 0 & 250 \\ 0 & 600 & 150 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 20 \\ 20 \\ 900 \\ 1 \end{pmatrix} \\
&= \frac{1}{900}\begin{pmatrix} 800 & 0 & 250 \\ 0 & 600 & 150 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 20 \\ 20 \\ 900 \end{pmatrix} \\
&= \frac{1}{900}\begin{pmatrix} 800 \times 20 + 0 + 250 \times 900 \\ 0 + 600 \times 20 + 150 \times 900 \\ 0 + 0 + 1 \times 600 \end{pmatrix} \\
&= \begin{pmatrix} 268 \\ 163 \\ 1 \end{pmatrix}
\end{aligned}
$$

Again the 268 is rounded up from 267.6667. Again either answer is ok.

Marking scheme

As above, but no new marks for the equation.

This is a question on image formation.