Mid-Level Vision: Multiview

1. Describe what is meant by the "correspondence problem" and briefly describe three scenarios which might require a solution to this problem.

The correspondence problem is the problem of finding the same 3D location in two or more images.

This problem might arise:

1. When using multiple cameras to obtain two, or more, images of the same scene from different locations - solving the correspondence problem would enable the 3D structure of the scene to be recovered.

2. When using a single camera to obtain two, or more, images of the same scene at different times - solving the correspondence problem would enable estimation of camera and object motion.

3. When comparing an image with one or more stored images - solving the correspondence problem would enable the similarity between the images to be determined and hence allow object recognition.

2. Briefly describe the methodology used in correlation-based and feature-based methods of solving the correspondence problem.

Correlation-based methods

Attempt to establish a correspondence for every pixel by matching image intensities in a window around each pixels

Feature-based methods

Attempt to establish a correspondence for a sparse sets of image locations, usually corners, using a feature description extracted from around each of those locations.

3. For feature-based methods of solving the correspondence problem, briefly explain what is meant by a "detector" and a "descriptor"

The detector is the method used to locate image features (or interest points) which are suitable for matching.

The descriptor is an array of feature values associated with each interest point. These descriptors are compared to determine which points match.

4. The two arrays below show the intensity values for each pixel in a stereo pair of 4 by 3 pixel images.

$$left: \begin{bmatrix} 4 & 7 & 6 & 7 \\ 3 & 4 & 5 & 4 \\ 8 & 7 & 6 & 8 \end{bmatrix} \qquad right: \begin{bmatrix} 7 & 6 & 7 & 5 \\ 4 & 5 & 4 & 5 \\ 7 & 6 & 8 & 7 \end{bmatrix}$$

Calculate the similarity of the pixel at coordinates (2,2) in the left image, to all pixel location in the right image, and hence, calculate the disparity at that point. Repeat this calculation for the pixel at coordinates (3,2) in the left image. Assume that (a) a 3 by 3 pixel window is used, (b) similarity is measured using the Sum of Absolute Differences (SAD), (c) the image is padded with zeros to allow calculation of similarity at the edges, (d) the cameras have coplanar image planes, (e) disparity is calculated as the translation from right to left.

For coplanar cameras, assuming the x-axes are also collinear, we can restrict the search for correspondence to the same row in the right image from which the pixel in the left comes. However, all similarity measures are shown here.

For point (2,2), SAD =

$$\begin{bmatrix} 36 & 33 & 30 & 39 \\ 15 & 12 & 9 & 24 \\ 36 & 34 & 35 & 42 \end{bmatrix}$$

Hence, best match is at location (3,2) in the right image.

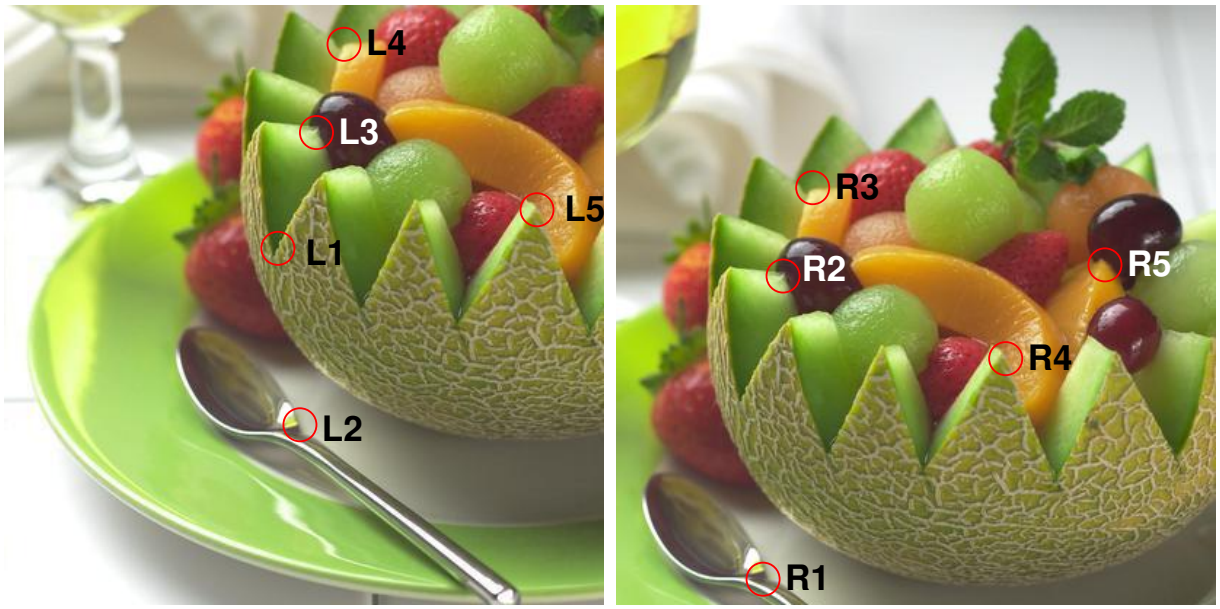
Disparity is left-right = (2,2)-(3,2) = (-1,0).

For point (3,2), SAD =

$$\begin{bmatrix} 40 & 35 & 32 & 39 \\ 25 & 0 & 11 & 22 \\ 40 & 36 & 35 & 42 \end{bmatrix}$$

Hence, best match is at location (2,2) in the right image.
Disparity is left-right = (3,2)-(2,2) = (1,0).

5. Below is a pair of images showing different views of the same scene.



The locations of interest points are indicated on each image, and vectors of features values for each of these interest point is given below:

Point	Feature Values	Point	Feature Values
L1	**(10, 4)**	**R1**	**(3, 7)**
L2	**(3, 8)**	**R2**	**(1, 1)**
L3	**(0, 2)**	**R3**	**(5, 7)**
L4	**(6, 9)**	**R4**	**(8, 0)**
L5	**(9, 1)**	**R5**	**(1, 2)**

For each interest point in the left image, find the best matching interest point in the right image assuming that similarity is measured using the sum of absolute differences (SAD).

For L1, SAD:
R1: 10; R2: 12; R3: 8; R4: 6; R5: 11
Therefore best match is R4.
For L2, SAD:
R1: 1; R2: 9; R3: 3; R4: 13; R5: 8
Therefore best match is R1
For L3, SAD:
R1: 8; R2: 2; R3: 10; R4: 10; R5: 1
Therefore best match is R5
For L4, SAD:
R1: 5; R2: 13; R3: 3; R4: 11; R5: 12
Therefore best match is R3
For L5, SAD:
R1: 12; R2: 8; R3: 6; R4: 2; R5: 9
Therefore best match is R4

6. The coordinates of the interest points in Question 5, are as follows:

Point	Coordinates	Point	Coordinates
L1	**(187, 168)**	**R1**	**(101, 394)**
L2	**(203, 290)**	**R2**	**(115, 186)**
L3	**(215, 87)**	**R3**	**(135, 128)**
L4	**(234, 28)**	**R4**	**(269, 243)**
L5	**(366, 142)**	**R5**	**(336, 178)**

Calculate the disparity at each point in the left image. Assume that (a) the cameras have coplanar image planes (although not collinear x-axes), (b) disparity is calculated as the translation from right to left.

Translation from R4 to L1 is (187,168)-(269, 243) = (-82, -75)
Translation from R1 to L2 is (203, 290)-(101, 394) = (102, -104)
Translation from R5 to L3 is (215, 87)-(336, 178) = (-121, -91)
Translation from R3 to L4 is (234, 28)-(135, 128) = (99, -100)
Translation from R4 to L5 is (366, 142)-(269, 243) = (97, -101)

7. Write pseudo-code for the RANSAC algorithm.

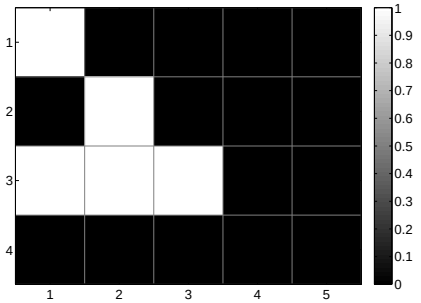
1. Randomly choose a minimal subset (a sample) of data points necessary to fit the model
2. Fit the model to this subset of data
3. Test all the other data points to determine if they are consistent with the fitted model (i.e. if they lie within a distance t of the model's prediction).
4. Count the number of inliers (the consensus set). Size of consensus set is model's support
5. Repeat from step 1 for N trials
After N trials select the model parameters with the highest support and re-estimate the model using all the points in this subset.

8. Apply the RANSAC algorithm to find the true correspondence between the two images in Question 5. Assume (a) that the images are related by a pure translation in the x-y plane, (b) that t (the threshold for comparing the model's prediction with the data) is 20 pixels, (c) 3 trials are performed and these samples are chosen in the order L1, L2, L3 rather than being randomly chosen.

Choose L1. Model is a translation of (-82, -75).
Locations of matching points predicted by this model are:
For L2; (203, 290)-(-82, -75) = (285, 365)
actual match is at (101, 394) hence this is an outlier for this model.
For L3; (215, 87)-(-82, -75) = (297, 162)
actual match is at (336, 178) hence this is an outlier for this model.
For L4; (234, 28)-(-82, -75) = (316, 103)
actual match is at (135, 128) hence this is an outlier for this model.
For L5; (366, 142)-(-82, -75) = (448, 217)
actual match is at (269, 243) hence this is an outlier for this model.
Hence, consensus set = 0.
Choose L2. Model is a translation of (102, -104).
Locations of matching points predicted by this model are:
For L1; (187, 168)-(102, -104) = (85, 272)
actual match is at (269, 243) hence this is an outlier for this model.
For L3; (215, 87)-(102, -104) = (113, 191)
actual match is at (336, 178) hence this is an outlier for this model.
For L4; (234, 28)-(102, -104) = (132, 132)
actual match is at (135, 128) hence this is an inlier for this model.
For L5; (366, 142)-(102, -104) = (264, 246)
actual match is at (269, 243) hence this is an inlier for this model.
Hence, consensus set = 2.
Choose L3. Model is a translation of (-121, -91).
Locations of matching points predicted by this model are:
For L1; (187, 168)-(-121, -91) = (308, 259)
actual match is at (269, 243) hence this is an outlier for this model.
For L2; (203, 290)-(-121, -91) = (324, 381)
actual match is at (101, 394) hence this is an outlier for this model.
For L4; (234, 28)-(-121, -91) = (355, 119)
actual match is at (135, 128) hence this is an outlier for this model.
For L5; (366, 142)-(-121, -91) = (487, 233)
actual match is at (269, 243) hence this is an outlier for this model.
Hence, consensus set = 0.

Therefore the true correspondence is given by the matches for L2, L4, and L5. The best estimation of the model is $\frac{1}{3}[(102,-104)+(99,-100)+(97,-101)] = (99.33,-101.67)$

9. Below is shown a simple 5 by 4 pixel binary image. The two arrays show the derivatives of the image intensities in the x and y directions.



$$I_x = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad I_y = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Given the x any y derivatives of the image intensities shown above, calculate the response of the Harris corner detector at each of the six central pixels, assuming (a) a value of k=0.05, (b) that products of derivatives are summed over an equally weighted, 3 by 3 pixel, window around each pixel.

$$R = \left[\sum I_x^2 \sum I_y^2 - (\sum I_x I_y)^2\right] - k\left[\sum I_x^2 + \sum I_y^2\right]^2$$

$$I_x^2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad I_y^2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad I_x I_y = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\sum I_x^2 = \begin{bmatrix} 3 & 3 & 1 & 0 & 0 \\ 3 & 4 & 2 & 1 & 0 \\ 2 & 3 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} \quad \sum I_y^2 = \begin{bmatrix} 3 & 4 & 2 & 1 & 0 \\ 5 & 7 & 4 & 2 & 0 \\ 3 & 5 & 3 & 2 & 0 \\ 2 & 3 & 2 & 1 & 0 \end{bmatrix} \quad \sum I_x I_y = \begin{bmatrix} 2 & 2 & 0 & 0 & 0 \\ 2 & 3 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

$$\sum I_x^2 \sum I_y^2 = \begin{bmatrix} 9 & 12 & 2 & 0 & 0 \\ 15 & 28 & 8 & 2 & 0 \\ 6 & 15 & 6 & 2 & 0 \\ 0 & 3 & 2 & 1 & 0 \end{bmatrix} \quad (\sum I_x I_y)^2 = \begin{bmatrix} 4 & 4 & 0 & 0 & 0 \\ 4 & 9 & 1 & 1 & 0 \\ 1 & 4 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

$$(\sum I_x^2 + \sum I_y^2)^2 = \begin{bmatrix} 36 & 49 & 9 & 1 & 0 \\ 64 & 121 & 36 & 9 & 0 \\ 25 & 64 & 25 & 9 & 0 \\ 4 & 16 & 9 & 4 & 0 \end{bmatrix}$$

$$R = \left[\sum I_x^2 \sum I_y^2 - (\sum I_x I_y)^2\right] - k\left[\sum I_x^2 + \sum I_y^2\right]^2$$

$$R = \begin{bmatrix} 9 & 12 & 2 & 0 & 0 \\ 15 & 28 & 8 & 2 & 0 \\ 6 & 15 & 6 & 2 & 0 \\ 0 & 3 & 2 & 1 & 0 \end{bmatrix} - \begin{bmatrix} 4 & 4 & 0 & 0 & 0 \\ 4 & 9 & 1 & 1 & 0 \\ 1 & 4 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} - 0.05 \times \begin{bmatrix} 36 & 49 & 9 & 1 & 0 \\ 64 & 121 & 36 & 9 & 0 \\ 25 & 64 & 25 & 9 & 0 \\ 4 & 16 & 9 & 4 & 0 \end{bmatrix}$$

$$R = \begin{bmatrix} 3.2 & 5.55 & 1.55 & -0.05 & 0 \\ 7.8 & 12.95 & 5.2 & 0.55 & 0 \\ 3.75 & 7.8 & 3.75 & 0.55 & 0 \\ -0.2 & 1.2 & 0.55 & -0.2 & 0 \end{bmatrix}$$

10. For the Harris corner detector, describe what type of image feature will give rise to the following values of R.
(a) R ≈ 0
(b) R < 0
(c) R > 0.

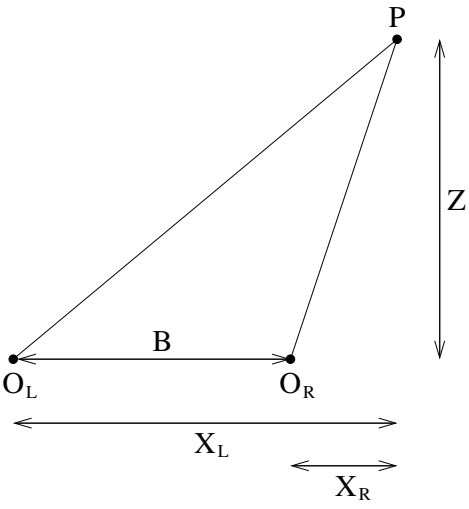
(a) $R \approx 0$ occurs where intensity values are unchanging
(b) $R < 0$ occurs at edges
(c) $R > 0$ occurs at corners.

11. Two cameras with identical focal lengths are set up so that their image planes are coplanar and their x-axes are collinear. Derive the relationship between the distance (Z) of a point from the cameras and the distance (B) that separates the origins of the two cameras.

For any camera, the image coordinates of a point (X,Y,Z) is

$$(u, v) = \left(\frac{fX}{Z}, \frac{fY}{Z}\right)$$

where all coordinates are relative to the cameras coordinate system.



Hence, for the left camera in the stereo pair:

$$(u_L, v_L) = \left(\frac{fX_L}{Z}, \frac{fY_L}{Z}\right)$$

For the right camera:

$$(u_R, v_R) = \left(\frac{fX_R}{Z}, \frac{fY_R}{Z}\right) = \left(\frac{f(X_L - B)}{Z}, \frac{fY_R}{Z}\right)$$

(using the coordinate system of the left camera). Hence:

$$u_L - u_R = \frac{fX_L}{Z} - \frac{f(X_L - B)}{Z} = \frac{fB}{Z}$$

$$Z = \frac{fB}{(u_L - u_R)}$$

12. Comment on the accuracy with which an object's depth can be measured with (a) changing distance, (b) changing baseline.

Accuracy will depend on the size of the disparity: the larger the disparity, the smaller the effects of small measurement errors.

$$d = u_L - u_R = \frac{fB}{Z}$$

(a) as the distance to the object increases, Z increases, and disparity reduces.
Hence, accuracy increases as the object comes closer.

(b) as the baseline, B, increases, the disparity increases.
Hence, accuracy increases as the baseline gets longer.

13. Briefly explain what is mean by the Epipolar constraint on the stereo correspondence problem.

The Epipolar constraint reduces the search for correspondence to a single line (the epipolar line) in the image.
Finding the epipolar line requires knowledge of the intrinsic and extrinsic parameters of the cameras.
For a simple coplanar configuration of cameras the epipolar lines are the horizontal scan lines of the images.

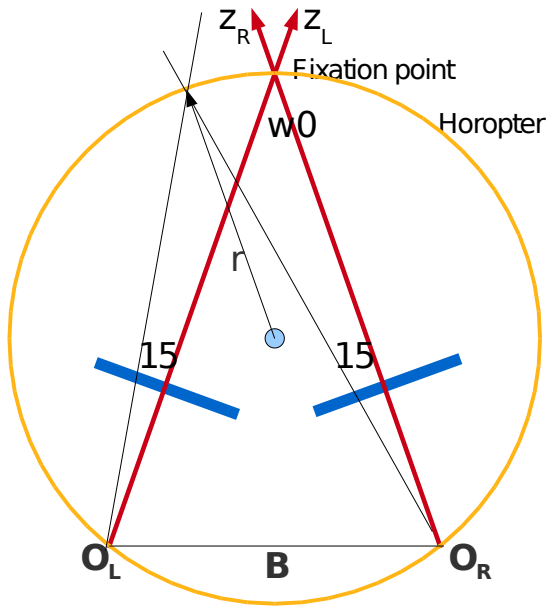
14. List other constraints applied to solving the stereo correspondence problem, and note circumstances in which they fail.

- *Maximum disparity (limit search to $\pm \frac{fB}{Z_{min}}$ around the point with zero disparity).*
- *Continuity (assume neighbouring points have similar disparities). Fails at discontinuities between surfaces at different depths.*
- *Uniqueness (assume each element in one image matches exactly one element in the other). Fails for surfaces angled such that they project different numbers of elements to each image, or where there is occlusion of a element in one image but not the other.*
- *Ordering (assume that matching elements occur in the same order along the conjugated epipolar lines). Fails where one surfaces is partially occluded by another.*

15. List constraints typically applied to solving the video correspondence problem, and note circumstances in which they fail.

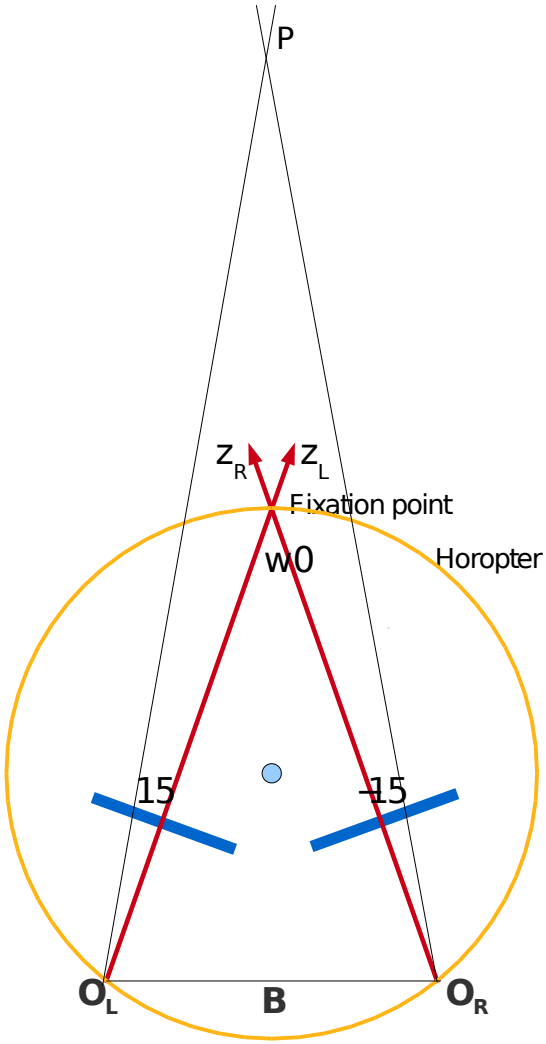
- *Spatial coherence (assume neighbouring points have similar optical flow). Fails at discontinuities between surfaces at different depths.*
- *Small motion (assume optical flow vectors have small magnitude). Fails if relative motion is fast or frame rate is slow.*

16. In a stereo vision system, the baseline between the camera centres is 400mm and the angle of convergence of the z-axes of the cameras is 60^o. Assume the z-axes of each camera make an equal angle with the baseline (*i.e.*, 60^o in this case). If the line-of-sight of a scene point makes angles α_L and α_R with the z-axes of the left and right cameras respectively, then what is the distance of the point from the horopter (a) $\alpha_L = \alpha_R = 15^o$, (b) $\alpha_L = +15^o$ and $\alpha_R = -15^o$, and (c) $\alpha_L = -15^o$ and $\alpha_R = +15^o$.



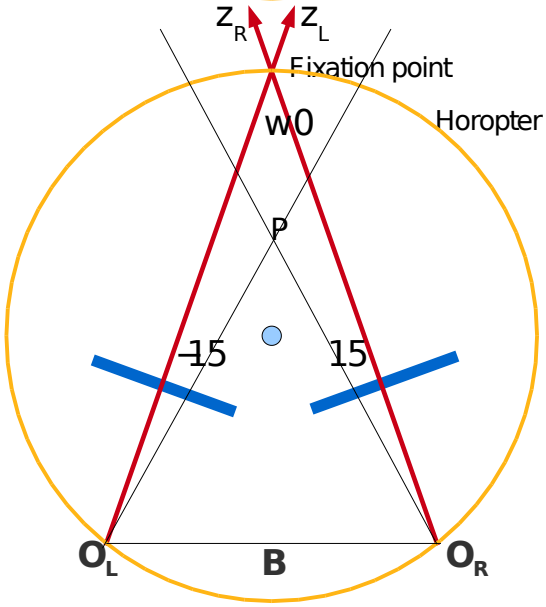
(a) $\alpha_L = \alpha_R = 15^o$
Disparity $= (\alpha_L - \alpha_R) = (15 - 15) = 0$

By definition, point is on the horopter, hence distance from the horopter is zero.



(b) $\alpha_L = +15^o$ *and* $\alpha_R = -15^o$
Disparity $= (\alpha_L - \alpha_R) = (15 - (-15)) = 30$

Angle $BO_RP = 75^o$
Hence, distance from the baseline to P is $0.5B\tan(75) = 746.4mm$
Distance of fixation point from baseline is $0.5B\tan(60) = 346.4mm$
Therefore distance between horopter and P is 746.4-346.4=400mm



(c) $\alpha_L = -15^o$ *and* $\alpha_R = +15^o$
Disparity $= (\alpha_L - \alpha_R) = (-15 - 15) = -30$

Angle $BO_RP = 45^o$
Hence, distance from the baseline to P is $0.5B\tan(45) = 200mm$
Distance of fixation point from baseline is $0.5B\tan(60) = 346.4mm$
Therefore distance between horopter and P is 200-346.4=-146.4mm

17. Give two reasons why the recovery of depth information is important for object recognition.

Depth **between objects** *provides a cue for segmentation and hence helps solve the problem of mid-level vision.*
Depth **within an object** *provides information about the shape of an object and hence helps solve the problem of high-level vision.*

18. Briefly describe two oculomotor cues to depth.

Accommodation *The shape of the lens in the eye, or the depth of the image plane in a camera, is related to the depth of objects that will be in focus. Hence, knowledge of these values provides information about the depth of the object being observed.*

Convergence *The rotation of eyes/cameras in a stereo vision system can vary to fixate objects at different depths. Hence, the angle of convergence provides information about the depth of the object being fixated.*

19. Briefly describe four monocular cues to depth.

Interposition *Nearer objects may occlude more distant objects. Hence occlusion (or interposition) provides information about relative depth.*

Size familiarity *Objects of known size provide depth information, since the smaller the image of the object the greater its depth.*

Texture gradients *For uniformly textured surfaces, the texture elements get smaller and more closely spaced with increasing depth.*

Linear perspective *lines that are parallel in the scene converge towards a vanishing point in the image. As the distance between the lines in the image decreases, so depth increases.*

Aerial perspective *Due to the scattering of light by particles in the atmosphere, distant objects look fuzzier and have lower luminance contrast and colour saturation.*

Shading *The distribution of light and shadow on objects provides a cue for depth.*

20. Briefly describe two motion induced cues to depth.

Motion parallax *As the camera move sideways, objects closer than the fixation point appear to move in a direction opposite to the camera, while objects further away appear to move in the same direction. The speed of movement increases with distance from the fixation point.*

Optic Flow *As a camera moves forward or backward, objects closer to the camera move more quickly across the image plane.*

Accretion and deletion *As a camera moves parts of an object can appear or disappear; these changes in occlusion provides information about relative depth.*

Structure from motion (kinetic depth) *Movement of an object or of the camera can generate different views of an object that can be combined to recover 3D structure.*

21. Define what is meant by the "aperture problem" and suggests how this problem can be overcome.

The aperture problem refers to the fact that the direction of motion of a small image patch can be ambiguous.

Particularly, for an edge information is only available about the motion perpendicular to the edge, while no information is available about the component of motion parallel to the edge.

Overcoming the aperture problem might be achieved by

1. integrating information from many local motion detectors / image patches, or

2. by giving preference to image locations where image structure provides unambiguous information about optic flow (e.g. corners).

22. Consider a traditional barber's pole as shown in this image:

When the pole rotates on its axis in which direction is the (a) motion field, (b) optic flow? How might this be explained by the aperture problem?

(a) the motion field is horizontal

(b) the optic flow is (roughly) vertical

The stripes of the pole provide ambiguous information. Only where the stripes meet the top and bottom and the sides of the pole are corners present and hence is there is unambiguous information.

If the visual system integrates information then there are more corners moving vertically, and a bigger component of vertical movement in the centre of the stripes. Hence, overall motion is seen as vertical.

If the visual system relies more heavily on unambiguous motion cues (i.e. corners), there are more corners at the sides of the pole moving vertically, than there are corners at the top and bottom of the pole moving horizontally. Hence, overall motion is seen as vertical.

23. Two frames in a video sequence were taken at times t and t+1s. The point (50,50,t) in the first image has been found to correspond to the point (25,50,t+1) in the second image. Given that the camera is moving at $0.1ms^{-1}$ along the camera x-axis, the focal length of the camera is 35mm, and the pixel size of the camera is 0.1mm/pixel, calculate the depth of the identified scene point.

The depth is given by: $Z = -\frac{fV_x}{\dot{x}}$.

The velocity of the image point is $\frac{25-50}{1} = -25$ *pixels/s.*

Given the pixel size this is equivalent to $0.0001 \times -25 = -0.0025$ *m/s.*

Hence, the depth is $Z = -\frac{0.035 \times 0.1}{-0.0025} = 1.4m$.

24. Two frames in a video sequence were taken at times t and t+1s. The point (50,70,t) in the first image has been found to correspond to the point (45,63,t+1) in the second image. Given that the camera is moving at $0.1ms^{-1}$ along the optical axis of the camera (*i.e.*, the z-axis), and the centre of the image is at pixel coordinates (100,140), calculate the depth of the identified scene point.

The depth is given by: $Z_2 = \frac{x_1 V_z}{\dot{x}}$.

The coordinates of the points with respect to the centre of the image are: (-50,-70,t) and (-55,-77,t+1).

The velocity of the image point is $\frac{-55-(-50)}{1} = -5$ *pixels/s.*

Hence, the depth is $Z_2 = \frac{-50 \times 0.1}{-5} = 1m$.

Alternatively, using the y-coordinates: $Z_2 = \frac{y_1 V_z}{\dot{y}}$.

The velocity of the image point is $\frac{-77-(-70)}{1} = -7$ *pixels/s.*

Hence, the depth is $Z_2 = \frac{-70 \times 0.1}{-7} = 1m$.

25. Give an equation for the time-to-collision of a camera and a scene point which does not require the recovery of the depth of the scene point.

Using this equation, calculate the time-to-collision of the camera and the scene point in the previous question, assuming the camera velocity remains constant.

time-to-collision $= \frac{x_1}{\dot{x}}$.

Hence, for the point in the previous question, time-to-collision $= \frac{-50}{-5} = 10s$.

(From the answer to the previous question, we know that the camera is 1m from the point and moving at $0.1ms^{-1}$*, so this agrees with the result above.)*

26. The arrays below show the pixel intensities in the same 1 by 5 pixel patch taken from four frames of a greyscale video. In order to segment any moving object from the background, calculate the result of performing (a) image differencing, (b) background subtraction. In both cases assume that the threshold is 50 and in (b) that the background is calculated using a moving average which is initialised to zero everywhere and which is updated using the formula $B(x,y) = (1-\beta)B(x,y) + \beta I(x,y,t)$ **where** $\beta = 0.5$**.**

$I(x,y,t1) = [190,200,90,110,90]$
$I(x,y,t2) = [110,170,160,70,70]$
$I(x,y,t3) = [100,60,170,200,90]$
$I(x,y,t4) = [90,100,100,190,190]$

(a) Image differencing.
$abs(I(x,y,t1) - I(x,y,t2)) = [80,30,70,40,20]$
$abs(I(x,y,t2) - I(x,y,t3)) = [10,110,10,130,20]$
$abs(I(x,y,t3) - I(x,y,t4)) = [10,40,70,10,100]$
applying threshold:
$abs(I(x,y,t1) - I(x,y,t2)) > 50 = [1,0,1,0,0]$
$abs(I(x,y,t2) - I(x,y,t3)) > 50 = [0,1,0,1,0]$
$abs(I(x,y,t3) - I(x,y,t4)) > 50 = [0,0,1,0,1]$
(b) Background subtraction.
at t1: $B = 0.5I(x,y,t1) = [95,100,45,55,45]$
$abs(I(x,y,t1) - B) = [95,100,45,55,45]$
at t2: $B = 0.5B + 0.5I(x,y,t2) = [102.5,135,102.5,62.5,57.5]$
$abs(I(x,y,t2) - B) = [7.5,35,57.5,7.5,12.5]$
at t3: $B = 0.5B + 0.5I(x,y,t3) = [101.25,97.5,136.25,131.25,73.75]$
$abs(I(x,y,t3) - B) = [1.25,37.5,33.75,68.75,16.25]$
at t4: $B = 0.5B + 0.5I(x,y,t4) = [95.625,98.75,118.125,160.625,131.875]$
$abs(I(x,y,t4) - B) = [5.625,1.25,18.125,29.375,58.125]$
applying threshold:
$abs(I(x,y,t1) - B) > 50 = [1,1,0,1,0]$
$abs(I(x,y,t2) - B) > 50 = [0,0,1,0,0]$
$abs(I(x,y,t3) - B) > 50 = [0,0,0,1,0]$
$abs(I(x,y,t4) - B) > 50 = [0,0,0,0,1]$