1. How can you use the Harris corner detector as an edge detector? How can you get the edge orientation?
   $\lambda_{\max}$ is high whenever there is an edge and can be used as an edge detector. $x_{\max}$ gives the orientation of the edge.

2. Each of the following filters is either a low-pass or a high pass filter. Say which is which:
   (a) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$
   (b) $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$
   (c) $\begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$
   (d) $\begin{bmatrix} -1 & 2 & -1 \end{bmatrix}$

   Low, Low, high, high

3. Suppose we are interested in image classification. In the table below, each column is a particular kind of feature vector, and each row is a particular transformation. In the cells of the table, write down if the feature vector on the column is invariant to the transformation on the row.

	Bag of words	SIFT	Image flattened as a vector	Image gradients flattened as a vector
Illumination changes				
Small translations				
Large translations				

	Bag of words	SIFT	Image flattened as a vector	Image gradients flattened as a vector
Illumination changes	Y	Y	N	Y
Small translations	Y	Y	N	N
Large translations	Y	N	N	N

4. Consider a bag-of-words based feature vector and a linear classifier. Recall that for a bag-of-words feature vector, we run k-means on SIFT feature vectors to get cluster centers that form our visual words. How does the k in this clustering step affect the dimensionality of the feature vector? How does increasing k affect overfitting?
   As k increases, the feature vector dimension increases, so the number of parameters in the linear classifier increase, so overfitting increases.

5. How do the epipoles relate to the fundamental matrix F? How do they relate to the locations of the camera?
   The epipoles are solutions to $Fx = 0$ and $F^T x = 0$. They are the image location of camera 1 in camera 2 and vice versa

6. Pixels at a larger depth have a ____________ disparity.
   Lower.

7. Object detection is often considered a harder problem than image classification. Write down 3 ways in which this is true.

   (a) The need to localize exactly.
   (b) The need to avoid duplicate detections.
   (c) The large number of candidate bounding boxes.

8. A plane in 3D is given by the equation $\mathbf{N}^T\mathbf{X} = d$. Under a rotation and translation, the point changes as $\mathbf{X}' = R\mathbf{X} + \mathbf{t}$. How does the equation of the plane change? How does the vanishing line change?
   $\mathbf{N}' = R\mathbf{N}, d' = d + \mathbf{N}^T R^T \mathbf{t}$, $\mathbf{N}'^T X = 0$.

9. Given a set of 3D points and the corresponding 2D image locations, we can get a set of linear equations in the entries of the camera projection matrix P of the form $A\mathbf{p} = 0$. What additional constraint can be imposed on $\mathbf{p}$ and why? How is this equation solved? How many equations do we need to solve for $\mathbf{p}$?
   We can say that $\|\mathbf{p}\| = 1$. This is because P operates on homogenous coordinates, so P can be freely scaled. This equation is solved by looking for the smallest eigenvector of A.

10. Given multiple images of the same scene with different light sources, which of the following can you estimate?

    (a) Albedo
    (b) Normals
    (c) Absolute depth
    (d) Relative depth

    1,2,4

11. Consider a pixel for which the BRDF $\rho(\theta_i, \theta_r) = \cos\theta_i$. Is the color of this pixel dependent on the viewing direction? What is the color of this pixel in terms of the lighting direction $\mathbf{L}$ and the surface normal $\mathbf{N}$?
    No. $(\mathbf{L}^T\mathbf{N})^2$.

12. Which of the following can prevent overfitting? Which can reduce training error?

    (a) Increasing the number of output channels for a convolutional layer.
    (b) Converting a convolutional layer to a fully connected layer.
    (c) Reducing the kernel size of a convolutional layer.

(d) Replacing a $k \times k$ convolutional layer (with identical number of input and output channels) with two convolutional layers: a $k \times 1$ and a $1 \times k$ respectively.

3 and 4 reduce parameters, so they likely reduce overfitting. 1 and 2 increase parameters, so they likely reduce underfitting, i.e., training error.

13. What is the computational cost of running a $k \times k$ convolutional layer with c_{out} filters on a collection of c_{in} feature maps of size $h \times w$? How many feature maps get produced as output?
$O(k^2 c_{out} c_{in} hw)$, c_{out}.

14. The R-CNN system uses segmentation to generate proposals. Why is this better than simply picking random bounding boxes?
Because segmentation is likely to identify groups of pixels that form a coherent object. Random bounding boxes might only cover part of an object, cover two objects or cover mostly background.