EN.520.665
HW1 Solution

Question 1

In Marr-Hildreth operator, Gaussian smoothing filter is used in Laplcaian of Gaussian (LoG) or the approximated version Difference of Gaussian (DoG). First, it can smooth the image and reduce the noise. The second order derivative of the smoothed image indicates the rate of change in the intensity gradient, i.e. captures local maxima in the gradient. Marr-Hildreth operator assumes the spatial coincidence assumption. If a zero-crossing segment is present in a set of independent $\nabla^2 G$ channels over a contiguous range of sizes and the segment has the same position and orientation in each channel, then the set of such zero-crossing segments may may be taken to indicate the presence of an intensity change in the image that is due to a single physical phenomenon.

Compared with Marr-Hildreth operator, Canny edge detector and SIFT also use Gaussian filter to smooth the image and reduce the noise. However, Canny edge detector uses the first derivative of the Gaussian operator. This can locate the horizontal and vertical gradients of the image use them to obtain gradient intensity and direction. Canny edge detector is based on step-edges corrupted by additive Gaussian noise.

SIFT also uses DoG similar to the Marr-Hildreth operator. However, instead of one particular Gaussian operator used in the Marr-Hildreth operator, SIFT uses Gaussian operators with different scale. SIFT takes the maxima/minima of the DoG images that occur at multiple scales. This is different to the Marr-Hildreth operator uses zero crossing to find the edges.

Question 2

We can extend the Marr-Hildreth operator to extract and keypoints in a sub-pixel accuracy. First, we apply DoG to the input image. Then the zero-crossing and the extrema of the DoG would be the edges and key points respectively. Now we can use curve fitting method on the image. For example, we can use the 3D curve fitting method in SIFT, which is done using the quadratic Taylor expansion of the Difference-of-Gaussian scale-space function, to locate more accurate edges and keypoints in a sub-pixel level.

Question 3

SIFT features are computed by first finding extrema in DoG across the scale space. The extrema are searched for over the entire scale space, and this step is designed so that the features are invariant to scale changes. SIFT features are insensitive to scale as they are normalized to unit length. This eliminates dependence on change in contrast where all pixels are multiplied by a constant. Additive changes in intensity due to changes in brightness do not affect SIFT features, as it does not affect the gradient values. SIFT features are thus insensitive to uniform additive and multiplicative changes in intensity values. By aggregating gradients in a small region in a histogram, these features are somewhat insensitive to changes in viewpoint.

Question 4

LBP feature at a location c, with P equally spaced pixels in a radius of R around the center pixel is given as follows:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{1}$$

where $s(x)$ is the sign function defined as:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{2}$$

This can be viewed as a binary pattern of length P. Rotating the image results in a shift in this binary pattern. Our goal in achieving a rotation invariant LBP is to define LBP such that this binary pattern does not change with rotation of the image. This can be achieved by taking a binary pattern and taking all the shifted versions of it, and defining LBP to be the minimum of these bit strings. This can be written as below:

$$LBP_{P,R}^{ri} = min(ROR(LBP_{P,R}, i)|i = 0, 1, ..., P - 1) \tag{3}$$

Here, $ROR(LBP_{P,R}, i)$ performs a right bitwise shift on $LBP_{P,R}$ i times. The minimum of all the shifted versions gives rotation invariant LBP. Some of these patterns correspond to different features in the image, such as edges, flat areas and dark spots, independent of the orientation.