The University Of Sheffield.

Autumn Semester 2012-2013

EEE6082 Computational Vision 4 MODEL ANSWERS

1. a. Entropy is used to define the feature space saliency.

A flattened/distributed histogram is more salient, because it results in a higher entropy value. **4**

b. (a) $H = -[\frac{1}{2} \times \log_2(\frac{1}{2}) + \frac{1}{2} \times \log_2(\frac{1}{2})] = 1$

(b) $H = -[\frac{1}{4} \times \log_2(\frac{1}{4}) + \frac{1}{4} \times \log_2(\frac{1}{4}) + \frac{1}{4} \times \log_2(\frac{1}{4}) + \frac{1}{4} \times \log_2(\frac{1}{4})] = 2$ **4**

c. A Gaussian function can be used to weight pixels in the centre more than those on the edge.

Partial volume/pixel estimation can be used to make discrete circular regions smoother. **6**

Parzen windowing or bin interpolation can be used to make the histogram less sparse.

d. First illustrate the locations and scales of the 4 regions of the original image in the transformed image. The coordinates of the centres of those 4 regions in the transformed image are (180, 20), (20, 20), (180, 180), (20, 180) and their scale (radius) all become 10 according to the scaling factor. Then it’s easy to see that two of the regions are corresponding to two detected regions in the transformed image with the overlap error less than 50%. So the repeatability rate is 2/4=50%. **6**

2. **a.** Keypoint matching between two images involve the following steps:

1. Find a set of distinctive keypoints in both images;

2. Define a region around each keypoint; **5**

3. Extract and normalise the region content;

4. Compute a local descriptor from the normalised region;

5. Match local descriptors.

b. For a detected interest point, gradually increase the scale of the region around the detected interest point and calculate the responses for different scales. Automatic scale selection is done by picking the maxima of the responses. **3**

c. The 4 histogram bins can be centred on four orientations as: 0/360, 90, 180, 270. For each bin, the counts are 7, 10, 16, 7, respectively, by taking into account interpolations between neighbouring bins. Make sure the counts are weighted by the gradient magnitudes. **6**

Other binnings of the orientation degrees are also acceptable if they are correctly weighted and interpolated.

d. **i)** SIFT is composed of the detection part and the description part. The description part of SIFT is similar to HOG with some differences in parameters.

ii) First compute the orientation histogram of the region, and select the dominant orientation. Then the region is rotated to a fixed orientation.

iii) First, two SIFT features are matched based on their similarity/distance; Second, two SIFT features can be matched based on the ratio of distances between the nearest neighbour and the second nearest neighbour. **6**

3. **a.** The Bag of Features (BoF) model for action recognition consists of the following steps:

   1. Extract spatio-temporal features from action sequences
   2. Learn 'visual vocabulary'
   3. Quantize features using 'visual vocabulary'
   4. Represent action sequences by frequencies of 'visual words'

   The 'visual vocabulary' is usually constructed using K-means clustering. **5**

   b. The feature points are (50, 60, 80, 100, 120, 150, 170)

   K = 3; The initial cluster centres can be (50, 100, 150) (Other cluster centres are also acceptable) **6**

   Allocate each feature point to the nearest cluster centre as follows:

   For cluster centre 50: 50, 60;

   For cluster centre 100: 80, 100, 120;

   For cluster centre 150: 150, 170.

   Recalculate the cluster centres for each cluster:

   The new centre for cluster 1 is: (50+60)/2 = 55;

   The new centre for cluster 2 is: (80+100+120)/3 = 100;

   The new centre for cluster 3 is: (150+170)/2 = 160.

   Allocate each feature point the a new cluster whose centre is nearest:

   For cluster centre 55: 50, 60;

   For cluster centre 100: 80, 100, 120;

   For cluster centre 150: 150, 170.

   It can be seen that the cluster centres do not change anymore.

   END

   c. An action sequence is represented as a histogram of the frequencies of the visual words.

   The dimension of the histogram is K as in K-means clustering. **4**

   d. MEI only indicates the location of motion;

   MHI indicates both the location and when that motion happens – brighter values in MHI correspond to more recent motion. **5**

   Advantages and disadvantages of BoF and motion template for action representation:

   BoF is based on local spatio-temporal local features, so it is more robust to occlusion and background changes, but less informative;

   Motion template is a holistic representation, so it is more informative but less robust to background changes and occlusion.

4. a. Face detection works as follows:

1. Scan the image with different window sizes;

2. For each location and window size, a binary classifier is applied to classify the current image patch into face or non-face (the binary classifier is learned beforehand). **6**

For face detection to work for both frontal faces and profile faces, two detectors should be learned using training examples of frontal faces and profile faces separately.

b. The three major components of this Viola-Jones face detector are:

1. Integral images for calculating Haar-like features;

2. AdaBoost for feature selection; **3**

3. Cascade of classifiers for speed.

c. $D-A=[ii(4)-ii(2)-ii(3)+ii(1)]-ii(1)=ii(4)-ii(2)-ii(3)$

$C+D-A-B=[ii(3)-ii(1)]+[ii(4)-ii(2)-ii(3)+ii(1)]-ii(1)-[ii(2)-ii(1)]$

$=ii(4)-2ii(2)$ **5**

d. i. (1) Calculate the distance between histograms; (2) Histogram intersection.

ii. Different colour images could have similar colour histograms, because histograms only encode the probabilities of colours but not the structural layout information.

iii. A simple improvement is to use a spatial pyramid and compute a histogram in each spatial bin. 6