The University Of Sheffield.

Autumn Semester 2011-2012

EEE6082 Computational Vision 4 MODEL ANSWERS

1. a. The properties of a good local region detector include:

- Invariance to global changes in imaging conditions, e.g. viewpoint, illumination
- Robustness to local perturbations, e.g. noise, motion
- Repeatability under intra-class variations
- Distinctiveness, i.e. detected regions are informative **4**

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

(b) $H = -[\frac{1}{2}\times\log_2(\frac{1}{2}) + \frac{1}{4}\times\log_2(\frac{1}{4}) + \frac{1}{4}\times\log_2(\frac{1}{4})] = 1.5$

(c) $H = -\{\frac{1}{4}\times\log_2(\frac{1}{4}) + 4\times[\frac{1}{8}\times\log_2(\frac{1}{8})] + 4\times[\frac{1}{16}\times\log_2(\frac{1}{16})]\} = 3$ **6**

c. First draw histograms of scales s, s-1 and s+1.

$$H = -[\frac{5}{9}\times\log_2(\frac{5}{9}) + \frac{4}{9}\times\log_2(\frac{4}{9})]$$

$$W = \frac{1}{2}\{\frac{9}{9-1}[(1-\frac{5}{9}) + (\frac{4}{9}-0)] + \frac{25}{25-9}[(\frac{5}{9}-\frac{1}{5}) + (\frac{4}{9}-\frac{4}{25}) + (\frac{12}{25}-0) + (\frac{4}{25}-0)]\}$$ **5**

d. First illustrate the locations and scales of the 4 regions of the original image in the transformed image. The locations of the top-left corners of those 4 regions in the transformed image are (20, 20), (20, 180), (180, 20), (180, 180) and their size is all 10x10. Then it’s easy to see that one of the regions is corresponding to one detected region in the transformed image with the overlap error less than 50%. So the repeatability rate is 1/4=25%. **5**

2. **a.** Parts-based – more robust against occlusion, overlaps, moving backgrounds, but less informative

Global – simpler, more accurate for small resolutions and "clean" data, but less robust to realistic data **4**

b. A sliding window traverses the whole image. For each location, the image window is classified by a binary classifier into object or non-object. The binary classifier is trained offline beforehand. **4**

c. The 6 histogram bins can be divided as (0-60), (60-120), (120-180), (180-240), (240-300), (300-360). For each bin, the counts are 5, 6, 0, 8, 0, 2, respectively. Make sure the counts are weighted by the gradient magnitudes.

Other divisions of the orientation degrees are also acceptable. **6**

d. **i)** The detection part and the description part

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**

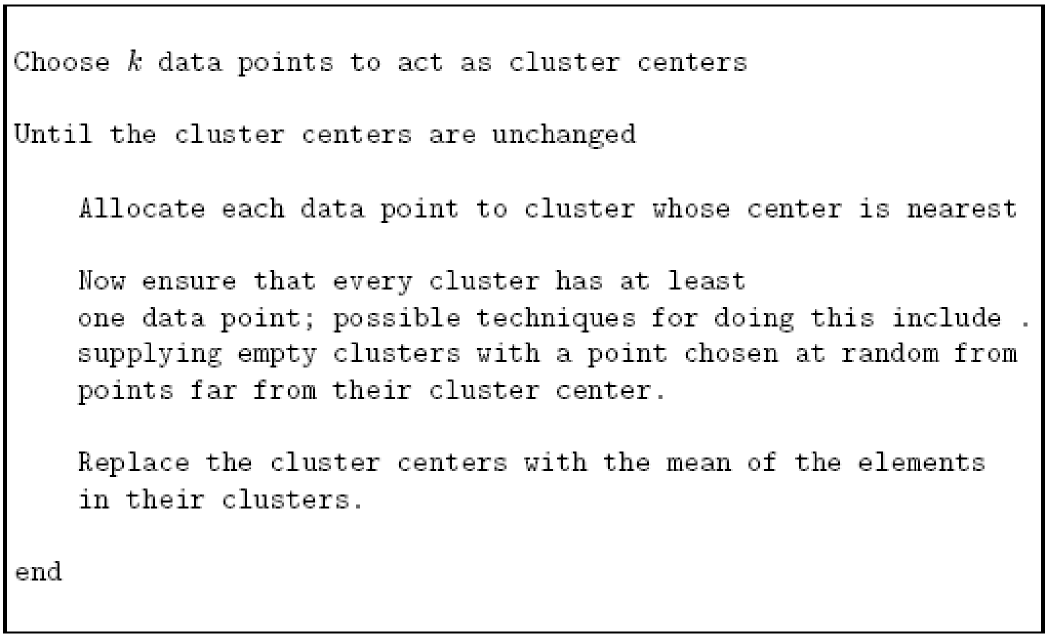
3. **a.** The Bag of Features (BoF) model consists of the following steps:

1. Extract features
2. Learn 'visual vocabulary'
3. Quantize features using 'visual vocabulary' **4**
4. Represent images by frequencies of 'visual words'

b. Vocabulary too small: visual words not representative for all image features

Too large: quantization errors, overfitting (training data will be too sparse to train system) **4**

c.



```
Choose $k$ data points to act as cluster centers

Until the cluster centers are unchanged

    Allocate each data point to cluster whose center is nearest

    Now ensure that every cluster has at least
    one data point; possible techniques for doing this include .
    supplying empty clusters with a point chosen at random from
    points far from their cluster center.

    Replace the cluster centers with the mean of the elements
    in their clusters.

end
```

5

Algorithm 16.5: *Clustering by K-Means*

d. i. Action: Atomic motion(s) that can be clearly distinguished (e.g. sitting down, running)

Activity: An activity contains several actions performed in succession (e.g. dining, meeting a person)

Event: A composite of activities (e.g. football match, traffic accident)

ii. BoF based action recognition is more robust to overlaps, occlusion, camera motion and background clutter, etc.

iii. The only difference is the features used. Object recognition uses 2D static features, and action recognition uses spatio-temporal motion features. **7**

4. **a.** Face detection tries to localize any face; face recognition attempts to distinguish one face from another.

Viewpoint/illumination change, facial expression, ageing, cosmetics/glasses/beard would make face recognition more difficult. **5**

b. Principal Component Analysis (PCA) and Fisher Linear Discriminant Analysis (LDA).

PCA is unsupervised and LDA is supervised. PCA preserves the maximum variance, and LDA maximizes scatter between classes and minimizes scatter within classes. **4**

c. $C - A = [ii(3) - ii(1)] - ii(1) = ii(3) - 2ii(1)$

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

$= 2ii(2) + 2ii(3) - 4ii(1) - ii(4)$ **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