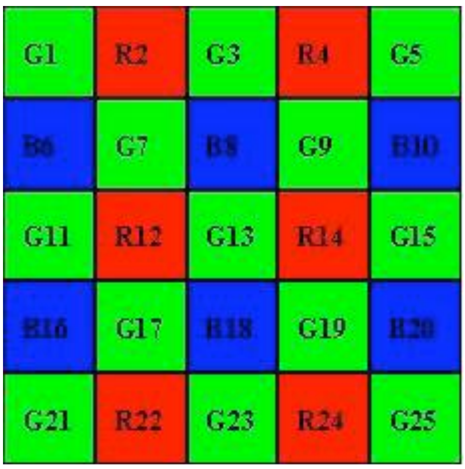
Autumn Semester 2013-2014

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) $H=-[\frac{1}{2}\times\log_2(\frac{1}{2})+\frac{1}{2}\times\log_2(\frac{1}{2})]=1$ (c) $H=-\{\frac{1}{2}\times\log_2(\frac{1}{2})+[\frac{1}{4}\times\log_2(\frac{1}{4})]+4\times[\frac{1}{16}\times\log_2(\frac{1}{16})]\}=2$	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 three of the regions are corresponding to three detected regions in the transformed image with the overlap error less than 50%. So the repeatability rate is 3/4=75%.	5

2. a.



Any answer describing the above pattern clearly is sufficient.

For example, the Bayer filter pattern subsamples RGB colours in one image plane and has 50% green, 25% red and 25% blue. The green colour is more densely sampled because human eyes are more sensitive to green. **4**

b. **i)** Aperture/shutter speed trade off to get the correct light exposure.

High aperture/fast shutter speed - captures fast moving objects without blurring, low depth of field, no sensitivity to camera shake. Reverse is true for low aperture/slow shutter speed.

First image has low aperture/fast shutter, second image has high aperture/slow shutter.

ii) High focal length/low field of view – weaker perspective, compression of depth, have to be further away from subject

To get weak perspective, stand far away from the object, use high focal length. To get strong perspective, stand close to object, use low focal length.

i) $\varphi = \tan^{-1}(\frac{d}{2f})$ **,** where *d* is the sensor size of the camera and *f* is the focal length.

10

c. The coordinates of point A should be $(-d\frac{x}{z}, \ -d\frac{y}{z})$, where (*x, y, z*) are the coordinates of B and *d*=1 is the distance between the origin and the projection plane PP. Therefore, the points are (10,0), (1.5,2), (-1,6/5).

6

3. **a.** Segmentation:

- provides a compact representation of scene

- allows us to determine which pixels lie on the same surface/find contingent regions/find boundaries of discrete objects **3**

- gives better semantic understanding of the scene

b. **i)**

```
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
```

Algorithm 16.5: *Clustering by K-Means*

ii) Intuitively, the optimal choice of K will strike a balance between maximum compression of the data using a single cluster, and maximum accuracy by assigning each data point to its own cluster. In general, experiment with different values for K on the same validation dataset, in order to discover the optimal choice of K.

iii) Reason is that initial cluster centres are randomly chosen, and that the final position of the cluster centres is dependent on this initial choice. In other words, the algorithm will find a “local optimum” for the cluster centres which is not guaranteed to be a “global optimum”. To mitigate this problem, usually clustering is done several times with different initial cluster centres in order to obtain the best result. **9**

c. **i)** Segmented regions are not effective for content-based image retrieval, because segmented regions are usually homogenous and contain little information.

ii) Salient regions are more informative/distinctive than segmented regions; salient regions are also more robust to various geometric and photometric transformations. **4**

d. Gaussian blurring of input image at various scales

Motivation is Gaussian blurring merges 'textured' pixel regions into 'averaged' solid colours, which are easy to segment. **4**

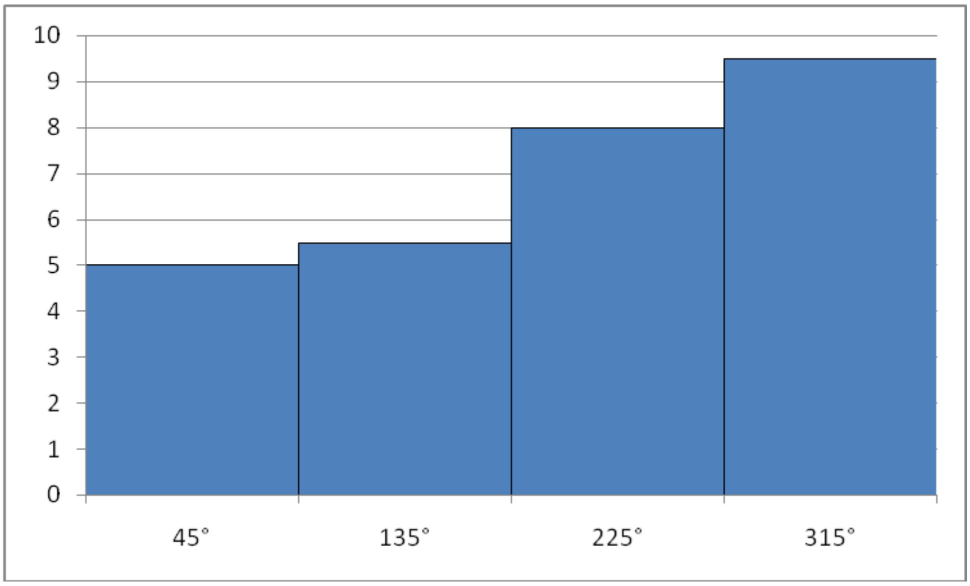
4. **a.** Parts-based – more robust against occlusion, overlaps, moving backgrounds

Global – simpler, more accurate for small resolutions and “clean” data **4**

b. Vocabulary too large – training data will be too sparse to train system.

Too small – features will not be sufficiently discriminative **3**

c.



Student should observe that some of the orientations lie exactly inbetween two bins, so the magnitude should be equally split between those two bins, as per interpolation. All other orientations lie in the centre of each bin, for simplicity. **7**

d. **i)** Face detection detects and localises any face regions; face recognition distinguishes faces from different people.

ii) A sliding window traverses the whole image. For each location, the image window is classified by a binary classifier into face or non-face. The binary classifier is trained offline beforehand using a large number of face regions and non-face regions. **6**