Solutions to the exam questions (June2, 2021)

Spatial statistics and image analysis (TMS016/MSA301)

1. (15p)

   a) Some image pre-processing is needed before detecting the wet areas. For example to reduce the noise in the image a Gaussian or a median filter can be used to smooth the image. Then, create the histogram of the data. If the histogram is bimodal a threshold τ can be manually chosen. For example, τ can chosen as the mean between the two modes. If it is hard to choose a threshold based on the histogram the image can be segmented into two classes using a Gaussian mixture model or the K-means algorithm. For example, the K-means algorithm is an unsupervised method consisting of the following steps:

      (a) Randomly select K observations as cluster centres.

      (b) Assign each observation to the closest cluster centre.

      (c) Compute the mean of each cluster and assign these as new cluster centres.

      (d) Repeat from step 2 until convergence

      This procedure is repeated a number of times with different starting cluster centres and the clustering that has the minimum total variation within the classes is preferred. (A description of the chosen model should be given here). After creating the binary image morphological operations might be used to regularize the image.

   b) At t=15 sweating is at an early stage and hence the majority of the wet areas have a circular shape, indicating that the neighboring wet areas have yet to merge together. Now, let A_s be the set of pixels classified as wet area for the sweat gland s. Then, the location for s can be estimated by the centroid of A_s with coordinates given by

$$\hat{x} = \frac{1}{\mid A_s \mid} \sum_{(x,y)\in A_s} x, \qquad \hat{y} = \frac{1}{\mid A_s \mid} \sum_{(x,y)\in A_s} y \qquad (1)$$

      where $\mid A_s \mid$ denote the area of A_s which can be estimated by the number of pixels in A_s.

      Those estimates will however be biased for the areas near the edges where the complete sweat areas are not observed. Also, this method will fail for the merged sweat areas at the lower right part of the image.

   c) Size features: The sweat spot areas $\mid A_s \mid$ can be estimated by counting the number of pixels in A_s. The perimeter of the sweat spot areas, $P(A_s)$ can be estimated by the number of pixels in A_s for which at least one of the eight neighboring pixels is not in A_s. The width of the sweat spot areas can be estimated by

$$W(A_s) = \max_{(x,y)\in A_s} x - \min_{(x,y)\in A_s} x \tag{2}$$

The height of the sweat spot areas can be estimated by

$$H(A_s) = \max_{(x,y)\in A_s} y - \min_{(x,y)\in A_s} y \tag{3}$$

Shape features: The compactness of A_s can be estimated by

$$C(A_s) = 4\pi \frac{\mid A_s \mid}{P(A_s)^2} \tag{4}$$

If A_s has perfect circular shape then $C(A_s) = 1$. The convexity of A_s can be estimated by

$$Co(A_s) = \frac{P(B)}{P(A_s)} \tag{5}$$

where B is the convex hull of A_s. If A_s is convex then $C(A_s) = 1$. The quotient between the height and width of a spot can also give us information about the shape. For a sweat spot areas that have yet to merge this feature is expected to be close to 1.

d) One possibility to classify the merged areas is to use the size or shape features introduced in (b) to cluster the data into 2 classes (merged areas or not merged areas). For example, using the sweat spot areas as the one of the features. Unsupervised algorithms for data clustering such as Gaussian mixtures models or K-means can be considered for this task. Note here that the parameter K needs to be tuned since $K = 2$ might not give us the optimal clustering.

2. (15p)

a) Pattern A: The F function of the pattern (blue) lies below the F function in the Poisson case and the G function lies above the G function in the Poisson case. These indicate that the pattern is clustered.

Pattern B: The F function (blue) lies below and the F function in the Poisson case indicating that the pattern is clustered. The G function equals to zero (and lies below the Poisson curve) up to the distance 0.04 and jumps to the value one after, no further information can be revealed from the G function plot. To summarize, it seems that the pattern is regular at short scale but there is also some clustering present.

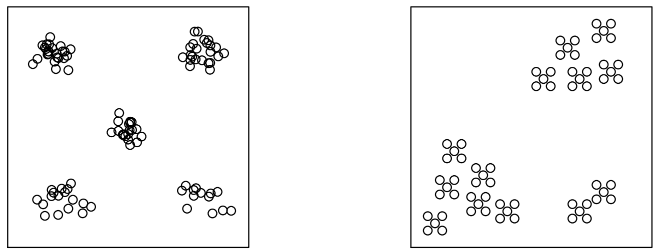
b) Pattern A: According the the $L(r) - r$ function, the pattern is clustered at small scale and regular at larger scale.

Pattern B: According the the $L(r)-r$ function, the pattern is regular at very small distances, then clustered.

c) Pattern A: Not really. F and G functions indicate only clustering (they see only the nearest neighbour) and do not see the long scale regularity that the centered L function shows.

Pattern B: F function does not see the short scale regularity and G function does not see the clustering. Combining the information from the F and G functions, we can determine that the pattern is regular at small distances and clustered in larger distances, and make the same conclusion as based on the $L(r) - r$ function. If we only saw one of the two curves, F or G, we would not be able to draw the same conclusion.

d) Pattern A may look like the pattern in the left and pattern B like the pattern in the right:



e) Pattern A: For example, a cluster process, where the parent points are from a hard-core process with a rather large hard-core radius. The daughter points can be generated e.g. as in the Matérn cluster process or Thomas process. The final pattern contains only the daughter points.
Pattern B: For example, a cluster process, where the parent points are from a Poisson process and the daughter points are generated as hard-core processes in a rather large disc around the parent points. The final pattern contains only the daughter points.