Section 4.

(1) Yes. Different samples' order may cause different cluster centers, the number of clusters will cause different clustering results.

(2) Yes. If the sample is an outlier, then the distance to its nearst token will greater than the threshold. (i.e. $dist(x_n, token_n) \geq h$), then the sample will become a new token, instead of being merged to other clusters.

(3) If h is a small number, then more samples will not be merged into existed clusters, in this case, these sample will become new cluster centers, and the total number of clusters is more then K-means.
If h is a big number, then most samples will be merged into pre-defined K clusters, in this case, all samples can be classified into K classes, which is the same number in K-means.
If token clustering and K-means have the same initial centers, i.e. the initial number of clusters are same, then token clustering will have at least K cluster in the end.
If the initial number of clusters in token clustering is less than K-means, then with a proper threshold, the number of clusters will increase and equal to K in the end.

(4.) In GMM, we set two components $N(x_n|\mu_1, \Sigma_1)$ and $N(x_n|\mu_2, \Sigma_2)$ to represent the distribution of pixels in hourses and background otherwise.
Then the model will be $p(x_n|z_k, \mu_k, \Sigma_k) = \sum_{k\in[1,2]} z_k N(x_n|\mu_k, \Sigma_k)$
After training (EM), when we predict a pixel x_i, if $N(x_i|\mu_1, \Sigma_1) > N(x_i|\mu_2, \Sigma_2)$ then its a pixel in hourse (foreground), otherwise it is in background.

(5) No. Some images may has different colors or different number of contents.

(6) Data augumentation for training set. For example, rotation flippling and clipping, to increase the amount of our training set.
Complex distributions. In this example, we can combine different kind of probability distribution models to increase the complexity of our model.