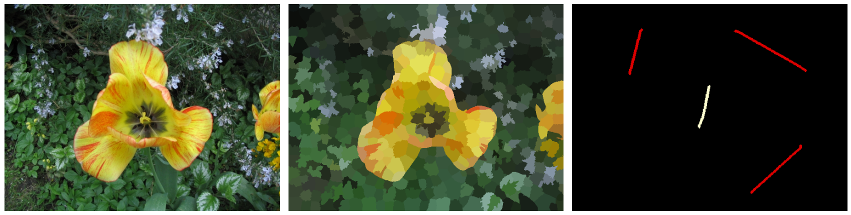
# Statistical Computer Vision (ECSE 626)
# Assignment 2

Due: March $25^{th}$, 2019 at 11:59PM

Attempt all parts of this assignment. The assignment will be graded out of total of **45 points**. Students are expected to write their own code. (Academic integrity guidelines can be found at `https://www.mcgill.ca/students/srr/academicrights/integrity`). Please submit your assignment solutions electronically via the mycourses assignment dropbox. Assignments received up to 24 hours late will be penalized by 30%. Assignments received more than 24 hours late will not be graded.

## Submission Instructions

1. Title two scripts as (i) **EM** (ii) **MRF**.
2. Reports should be submitted in the form of PDF **or** Jupyter notebook. **Other formats will not be accepted.**
3. Comment your code appropriately.
4. Do not submit the provided input images. Assume that the image folder and the face dataset are kept in a same directory as the code. Make sure that the submitted code runs without error. Add a README file if required.
5. Do not submit the output images separately. Include the output images within your report.
6. If external libraries were used in your code, please specify their name and version in the README file.
7. Answers to reasoning questions should be comprehensive but concise.
8. Submissions that do not follow the format will be penalized by 10%.



(a) (b) (c)

Figure 1: (a) Image to be segmented (b) an example of superpixel image (c) scribble mask depicting areas that are in the background (red) and foreground (white) [1].

# 1 Introduction

In this assignment, you will gain practice with unsupervised and supervised image segmentation. You will explore both EM (unsupervised) and Bayesian MRF methods (supervised). All methods should use the image shown in Figure 1(a).

# 2 Expectation Maximization (EM)

## 2.1 Fit Gaussian Mixture Model (GMM)

Consider a color image where, at each pixel $i$, we have observations $x_i = [R_i, G_i, B_i]^T$. Let the observation matrix be $X = [x_1, x_2, \ldots, x_N]$, where $N$ is the total number of pixels. You will now explore the process of fitting a 3-D GMM on $X$ using the EM algorithm. For segmentation, the value at every pixel is derived from the index of the Gaussian component with maximum log likelihood. Let $m$ be the number of Gaussian components needed to segment the image into $m$ classes. At every pixel, $i$, predict the segmentation class $f_i \in \{1, 2, \ldots, m\}$ for $m \in \{2, 3, \ldots 8\}$. Show the segmentation results for all cases. **(10 points)**

## 2.2 Finding the optimal value of $m$ (Model Selection)

Supposing the value of $m$ used in segmenting the given image is unknown. In this section, you will explore how the BIC can be used to find the optimal value of $m$. The BIC is a general Bayesian approach used in model selection. It relates to the Akaike information criterion and the minimum description length (see MacKay, p. 352). The BIC penalizes complexity in a model, where complexity refers to the number of parameters in the model. For example, consider fitting polynomials of different degrees. Higher-degree polynomials will yield a lower error but is more likely to overfit.

Supposing $n$ is the total number of pixels in the image, $\theta$ is the estimated model parameters (in this case, the GMM parameters), and $m$ is the number of parameters in a model. The BIC can be estimated using the following approximation:

$$BIC \approx -2 \sum_{x_i \in X} ln(p(x_i|\theta)) + m \ln(n)$$

Compute the BIC for models with $m \in \{2, 3, \ldots 8\}$. Repeat this 5 times with different random initializations of $\theta$. Plot the average BIC as a function of $m$. What is the optimal number of Gaussian components according to your plot? **(5 points)**

# 3 Bayesian Markov Random Field (MRF)

You will now explore developing a supervised image segmentation framework based on a Bayesian MRF (see Figure 1(b)). First, you will segment image $X$ into superpixels. Superpixels group neighbouring pixels with similar intensities, colors, etc. Denote the resultant image $S$, where each superpixel $i$ has value $s_i$, the mean $RGB$ color of the member pixels. (Note: You can generate a superpixel image using MATLAB function `superpixels` or Python function `skimage.segmentation.slic` or any other equivalent function.) **(1 point)**

The supervision for the segmentation will be derived from a scribble mask $M$ shown in Figure 1(c). $M$ should be interpreted as follows: If pixel $m_i \in M$ is yellow then the corresponding parent superpixel $i$ belongs to the $foreground$ and if $m_i$ is red then the corresponding parent superpixel $i$ belongs to the $background$. All the ground truth foreground superpixel values $s_i$ are grouped together into a matrix $S_{fg}$ (similarly for background superpixels, we have $S_{bg}$). **(2 points)** With these examples of ground truth values, you will now predict the labels $f_i \in \{foreground, background\}$ for the remaining superpixels $i \in S$.

Let $F$ be the joint label configuration of $\{f_i\}$. The posterior density function becomes:

$$p(F|S) \propto p(S|F)p(F)$$

## 3.1 Likelihood: $p(S|F)$

Estimate two Gaussian distributions, $\mathcal{N}_{fg}(\mu_{fg}, \Sigma_{fg})$ and $\mathcal{N}_{bg}(\mu_{bg}, \Sigma_{bg})$ for $S_{bg}$ and $S_{fg}$ respectively. Report the mean and covariance of $\mathcal{N}_{fg}$ and $\mathcal{N}_{bg}$. The likelihood models for superpixels in $S$ is defined as follows **(3 points)**:

$$\begin{aligned}
P(s_i|f_i = foreground) &= \mathcal{N}_{fg}(s_i; \mu_{fg}, \Sigma_{fg}) \\
P(s_i|f_i = background) &= \mathcal{N}_{bg}(s_i; \mu_{bg}, \Sigma_{bg}) \\
p(S|F) &= \prod_i P(s_i|f_i) \quad (\because \text{conditional Independence})
\end{aligned}$$

## 3.2 Prior: $p(F)$

Build a binary adjacency matrix $A$ for $S$ as follows. $A$ is a matrix of size $N \times N$ if there are $N$ superpixels. $A(i,j) = 1$ if superpixels $i$ and $j$ share a boundary and $A(i,j) = 0$ otherwise. Show $A$ in your report. **(3 points)**

Let $E = \{(i,j)|A(i,j) = 1\}$. Here, we can define the prior as follows:

$$\begin{aligned} P(f_i, f_j) &= exp\{-\beta \times I(f_i \neq f_j)\} \\ p(F) &= \prod_{(i,j)\in E} P(f_i, f_j) \end{aligned} \tag{1}$$

where $I$ is the identity matrix. $\beta$ is a hyperparameter to be tuned (explorations of the value of $\beta$ will be performed in section 3.3.3) **(2 points)**.

## 3.3 Posterior

Recall that MAP solution of the MRF can be found using various methods such as Metropolis sampling, Gibbs sampling, Marginal Posterior Mode (MPM), and Iterated Conditional Mode (ICM). In this assignment, you will explore using Metropolis sampling, a MCMC algorithm. Begin at time $t = 0$, by sampling label configuration $F^{(0)}$ from a distribution $\pi_\mathbf{0}(F)$. Progress by sampling $F^{(t)}$ from the proposal distribution $q(F^{(t)}|F^{(t-1)})$ and accepting $F^{(t)}$ with probability $\lambda = min\big(1, \frac{P(F^{(t)},S)}{P(F^{(t-1)},S)}\big)$ at every time $t$. Definitions of $\pi_\mathbf{0}$ and $q(F^{(t)}|F^{(t-1)})$ are given below.

### 3.3.1 Distribution $\pi_\mathbf{0}$

Consider the GMM with $m = 2$ that you estimated in section 2.1 as $\pi_0$. Initialize the label configuration $F$ with the results of the corresponding segmentation predictions estimated in Section 2.1 **(1 point)**.

### 3.3.2 Distribution $q(F^{(t)}|F^{(t-1)})$

Let proposal distribution be Gaussian. The variance of this Gaussian is a function of time $t$ **(1 point)**.

$$\begin{aligned} q(F^{(t)}|F^{(t-1)}) &= \prod_i q(f_i^{(t)}|f_i^{(t-1)}) \qquad (\because \text{ conditional independence}) \\ q(f_i^{(t)}|f_i^{(t-1)}) &= \mathcal{N}(0, \sigma^2(t)) \end{aligned}$$

Let $\sigma(t)$ be defined as:

$$\sigma(t) = \sqrt{\frac{3}{log(1+t)}}.$$

#### 3.3.3 Visualization

- Run Metropolis sampling for a sufficiently large number of times **(10 points)**.
- Visualize the progress by showing $F^{(t)}$ along with $q(f_i^{(t)}|f_i^{(t-1)})$ at five different intermediate timestamps **2 points**.
- At the end, sample 10 instances of $F^{(t)}$ and compute pixel-wise mode. Display the resulting segmentations **(2 points)**.
- Repeat this exercise for three different values of $\beta$ (from equation 1) of your choice. Report the effect of varying $\beta$ **(3 points)**.

## References


[1] V. Gulshan, C. Rother, A. Criminisi, A. Blake and A. Zisserman. Geodesic star convexity for interactive image segmentation. Proceedings of Conference on Vision and Pattern Recognition (CVPR 2010).