Accurate Image Segmentation using Gaussian Mixture Model with Saliency Map

Saksham Rathi and Navya

GMM and its limitations

- GMM (Gaussian Mixture Model) is a tool used for image segmentation and image classification.
- One main limitation of GMM is that it does not consider spatial information into account.
- This is because it assumes independence between different dimensions of input.
- So, it will perform poorly on datasets having spatial relationships in its points.
- Some other limitations include slow convergence and sensitivity to initialization.
- A common way to handle neighbouring pixels dependencies is the use of Markov Random Fields (MRF), but the use of MRF is computationally expensive.

Introduction

- When the goal of an application is the object recognition of an image, a visual saliency map is constructed by the combination of multi-scale low level image features, such as intensities, colours and orientations.
- We will use a saliency map to incorporate context-based spatial information into the conventional GMM for image segmentation.
- Our model known as GMM-SMSI (GMM with spatial information extracted from saliency map) is divided into two main steps.
- Firstly, a saliency map detection is obtained by means of the image spectral residual.
- Secondly, the saliency map is incorporated as spatial information into the conventional GMM.

Saliency Map

- Based on human visual system. Developed for image understanding and object recognition.
- The basic use of saliency is to suppress the response of frequently occurring features and to keep abnormal features.
- The average spectrum A(f) of an image can be approximated by filtering the image log spectrum L(f) by a local average filter h_n(f)

$$\mathcal{A}(f) = h_n(f) \times \mathcal{L}(f),$$

- The spectral residual R(f) is given by:

$$\mathcal{R}(f) = \mathcal{L}(f) - \mathcal{A}(f).$$

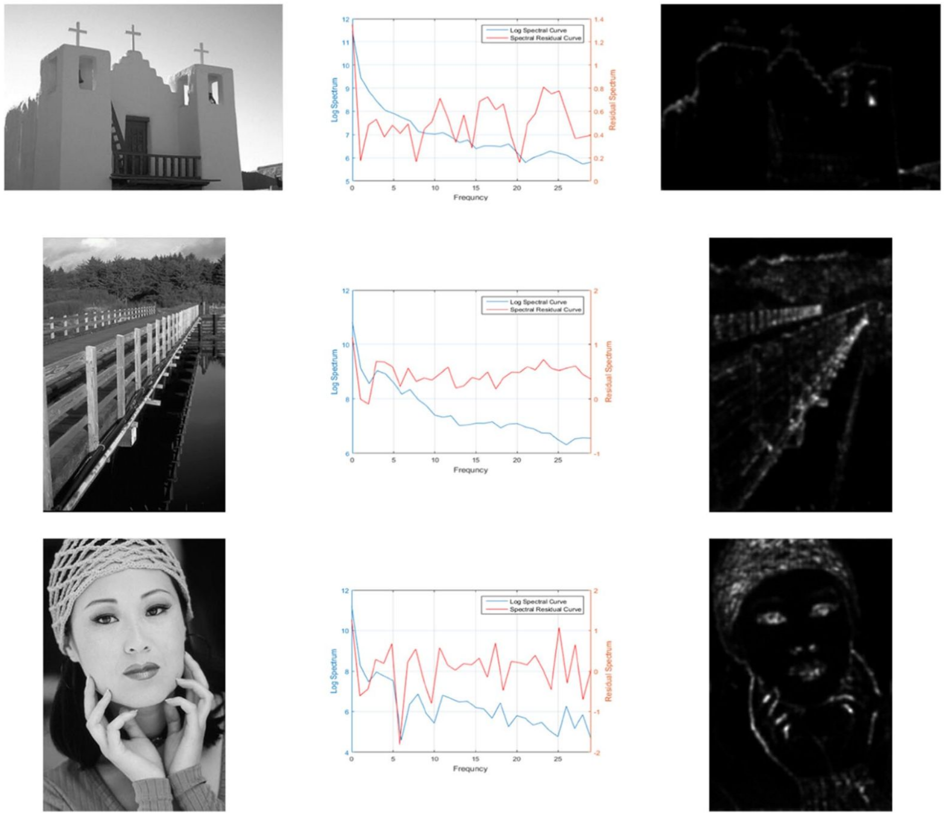
- The Saliency map is then calculated as follows: (where G(x) is the gaussian filter and F-1 is the inverse fourier transform.

$$S(x) = \mathcal{G}(x) * \mathcal{F}^{-1}\{\exp[\mathcal{R}(f) + \mathcal{P}(f)]\}^2,$$

Input images

Log spectrum curve and Spectral residual curve

Salience map



Conventional GMM

- For conventional GMM, the following two equations hold:

$$\Phi(x_i|\theta_{\mathrm{j}}) = \frac{1}{2\pi|\Sigma_j|}\exp\left[-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_j)\right],$$

$$f(x_i|\Pi, \Theta) = \sum_{j=1}^{L} \pi_j \Phi(x_i|\theta_j),$$

Saliency-weighted GMM

- The saliency-weighted GMM is given as:

$$f(x_i|\Psi) = \sum_{j=1}^{L} \pi_{ij} \left[\sum_{m \in \mathcal{N}_i} \frac{S(x_m)}{R_i} p(x_m|\Theta_j) \right],$$

where π_{ij} denotes the probability that the pixel x_i belongs to class j, π_{ij} satisfies the constraints $\pi_{ij} \geq 0$ and $\sum_{j=1}^{L} \pi_{ij} = 1$; $\mathcal{N}_i$ denotes the neighborhood of the pixel x_i; R_i is the sum of the saliency map values inside $\mathcal{N}_i$; $\mathbf{\Psi}$ denotes the parameters set containing all the parameters $\mathbf{\Psi} = \{\pi_{11}, \pi_{12}, \ldots, \pi_{1L}, \pi_{21}, \pi_{22}, \ldots, \pi_{2L}, \pi_{N1}, \pi_{N2}, \ldots, \pi_{NL}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_L\}$; and $S(x_m)$ is the saliency map value at location x_m.

- On applying EM algorithm for the parameters estimation in our model:

$$Q = \sum_{i=1}^{N} \sum_{j=1}^{L} \gamma_{ij} \left[\sum_{m \in \mathcal{N}_i} \frac{S(x_m)}{R_i} \log p(x_m | \tau_j) + \log \pi_{ij} \right].$$

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} \sum_{m \in \mathcal{N}_i} \gamma_{ij}^{(t)} \frac{S(x_m)}{R_i} x_m}{\sum_{i=1}^{N} \gamma_{ij}^{(t)}};$$

$$\gamma_{ij}^{(t)} = \frac{\pi_{ij}^{(t)} \sum_{m \in \mathcal{N}_i} \frac{S(x_m)}{R_i} p\left(x_m | \theta_j^{(t)}\right)}{\sum_{h=1}^{L} \pi_{ih}^{(t)} \sum_{m \in \mathcal{N}_i} \frac{S(x_m)}{R_i} p\left(x_m | \theta_h^{(t)}\right)}.$$

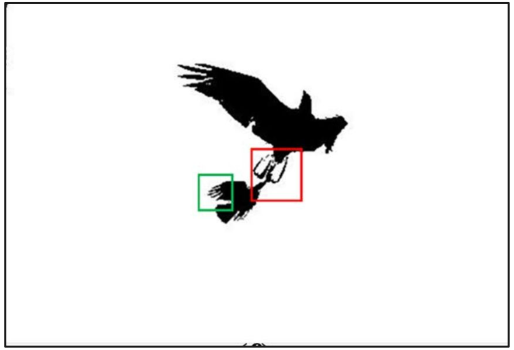
$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{N} \sum_{m \in \mathcal{N}_i} \gamma_{ij}^{(t)} \frac{S(x_m)}{R_i} (x_m - \mu_j^{(t)})(x_m - \mu_j^{(t)})^T}{\sum_{i=1}^{N} \gamma_{ij}^{(t)}}.$$

$$\pi_{ij}^{(t+1)} = \frac{\sum_{m \in \mathcal{N}_i} S(x_m) \gamma_{mj}^{(t)}}{\sum_{h=1}^{L} \sum_{m \in \mathcal{N}_i} S(x_m) \gamma_{mh}^{(t)}}.$$

EXPERIMENTS

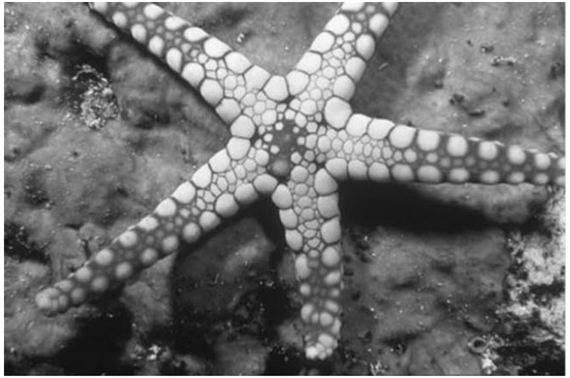
Tiny Object in Large Background Region

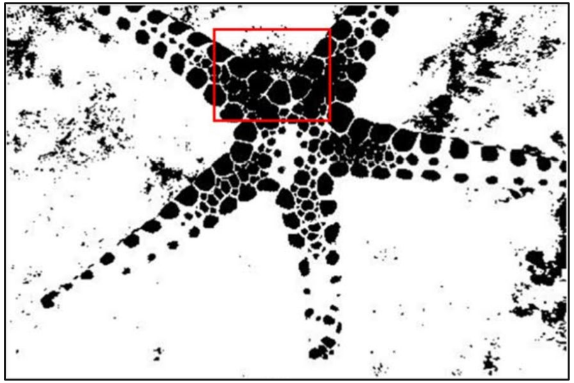
- GMM-SMSI with a probability rand (PR) index = 0.9864 was able to distinguish the two birds.
- The wings of the little bird (green square) also showed more details compared to other methods.



Large Object in Small Background Region

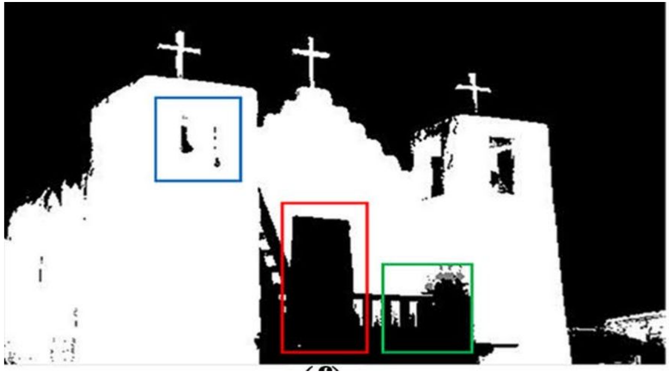
- An image with a relatively large starfish in the seabed was chosen for the second experiment.
- The algorithm separated the large starfish and the background clearly.
- It also obtained a high PR index value.





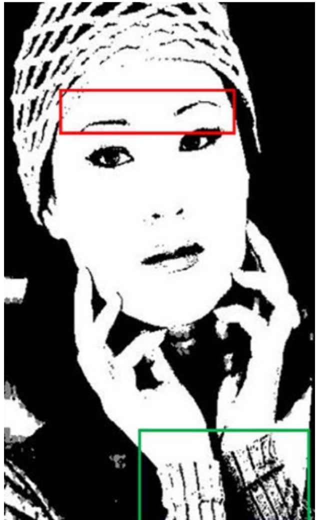
Building

- The segmentation showed more details in the church such as stairs, the left window and the door.



Human Face

- GMM SMSI showed more details of the human face, such as the full eyebrow information. It was also true when considering the texture of the clothes.



Conclusion

- This algorithm had better mean of PR indices for all images of the dataset.
- The computation time is also low because the spatial information is computed only once and this information helps the EM algorithm to converge faster.
- Moreover, the saliency map extraction is independent of GMM, which makes the proposed model easy to implement.
- **In summary, the proposed GMM-SMSI is an accurate, robust and fast algorithm which can be easily implemented and has a good execution time performance.**