The mean shift procedure is an extremely versatile tool for feature space analysis and can provide reliable solutions for many computer vision tasks. It exploits the fact that the most significant information in an image is located at the modes of the distribution of the image pixels. Using the gradient descent procedure we try to push most image pixels to their local model values to segment or cluster the image. A segmented image holds just as much information as the original image and the computer vision tasks become less computationally expensive.

The working of the procedure

In meanshift we locate the modes without estimating the ~~kernel~~ density function.

→ n data points a_i $i = 1$ to n of d-dimensions

$$f(x) = \frac{1}{n} \sum_{i=1}^{n} K_H (x - a_i)$$

K_H = kernel density functn

$$K_H = |H|^{-1/2} \, k(H^{-1/2} x)$$

and the kernel satisfies

$$\int_{R_d} k(x) \cdot dx = 1$$ as it's just an the prob density function.

→ We generate this multivariate kernel from a univariate kernel

$$K^S(x) = a_{k,d} \, k_1(\|x\|)$$

when $K^S(x)$ is a radially symmetric kernel

We are only interested in a class of radially symmetric kernels satifying $k(x) = c_{k,d} \, k(\|x\|^2)$ — (2)

$c_{k,d}$ makes it integrate to 1.

=> The bandwidth matrix H is taken to be diagonal

$$H = \mathrm{diag}\left[h_1^2 \cdots h_d^2\right]$$

$$H = h^2 I$$

By using only 1 bandwidth kernel, our kernel density estimator becomes -

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad - (1)$$

The Epanechnikov Kernel or E-kernel is a suitable kernel as it best minimizes the error in density estimation in our case.

In multivariate, the kernel become

$$K_N(x) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}\|x\|^2\right)$$

Using (1) and (2), our density estimator becomes

$$f_{h,k} = \frac{C_{k,d}}{nh^d} \sum_{i=1}^{n} k\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \quad - (4)$$

The modes are located at $\nabla f(x) = 0$

$$\star \quad \nabla f_{h,K} = \frac{2\, c_{k,d}}{n h^{d+2}} \sum_{i=1}^{n} (x - x_i)\, k'\left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \qquad -(3)$$

We define $g(x) = -k'(x)$

Lets define a kernel $G(x) = c_{g,d}\, g(\|x\|^2)$

$c_{g,d}$ = normalisation constant

$\Rightarrow$ $k(x)$ is called the shadow of kernel $G(x)$. $[g = k']$

Substituting $g(x)$ into (3) :-

$$\star \quad \nabla f_{h,K}(x) = \frac{2\, c_{k,d}}{n h^{d+2}} \sum_{i=1}^{n} (x_i - x)\, g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)$$

We can re-write it as

$$= \underbrace{\frac{2\, c_{k,d}}{n h^{d+2}} \left[\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \right]}_{\text{Proportional to density estimate at } x \text{ computed with Kernel } G} \left[\frac{\sum_{i=1}^{n} x_i\, g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right]$$

from (4)

$$f_{h,G} = \frac{C_{g,d}}{nh^d} \sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)$$

The second term is the mean shift

$$m_{h,G}(x) = \frac{\sum_{i=1}^{n} x_i\, g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x$$

Now we can write $\nabla f_{h,K}$ as

$$\nabla f_{h,K} = f_{h,G}(x)\, \frac{2\, C_{k,d}}{h^2\, C_{g,d}}\, m_{h,G}(x)$$

$$m_{h,G}(x) = \frac{1}{2} h^2 c \, \frac{\nabla f_{h,K}(x)}{f_{h,G}(x)}$$

$\Rightarrow$ It can be seen that the mean shif vector computed with kernel G is ~~proportinal~~ proportional to the normalised

In Summary

gradient density estimate obtained with kernel K.

=) mean shift vector always points toward greatest increase in density, therefore it can define a path to the stationary point of estimated density.

=) Upon successive computation of the mean shift vector $m_{h,G}(x)$ we are guaranteed to converge at a point where the gradient estimate is 0.

=) regions where density is low, are of no importance and in such regions mean shift steps are large. ~~a~~ Near local maxima, the steps are small and analysis is more refined.

My Procedure

We can establish a sufficient condition for convergence.

$$y_{j+1} = \frac{\sum_{i=1}^{n} x_i \, g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} \qquad j=1,2,..$$

Mean Shift Filtering -
We perform mean shift filtering using the condition of convergence established above.

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} \qquad j = 1, 2, \ldots$$

For every pixel in the image we perform the above operation for a number of iterations. Its been restricted to 2 to ease the computational requirements.

Steps to compute each pixel's value
1. Initialise j=1, y(i,1) = x(i) . i is the current pixel.
2. We compute y(i,2),y(i,3) .. for the number of iterations we desire
3. Finally we assign the updated pixel values to a new image

Mean Shift Segmentation -
Image segmentation is decomposition of a gray level or color image into homogeneous tiles and is arguably the most important low-level vision task. We perform segmentation after running the mean shift filtering procedure on the image.
Steps
1. Run mean shift filtering for a number of iterations and compute the image
2. Take the (hr,hs,m) i..e the range and spatial bandwidth and the number of clusters
3. Use K-means clustering technique by considering the norm of (Hr,Hs) as the distance metric. Hr is the distance of a point from the cluster centroid in the range domain and Hs is the distance of the point from the cluster in the spatial domain
4. For each pixel i assign the range value of its cluster centroid.

Results

We can see that the mean shift filtering has smoothened the image pretty well without losing out on much of the information. Segmentation has also produced results w.r.t its parameters.