World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:1, No:8, 2007

Segmentation of Images through Clustering to Extract Color Features: An Application for Image Retrieval

M. V. Sudhamani, and C. R. Venugopal

Abstract**—**This paper deals with the application for content-based image retrieval to extract color feature from natural images stored in the image database by segmenting the image through clustering. We employ a class of nonparametric techniques in which the data points are regarded as samples from an unknown probability density. Explicit computation of the density is avoided by using the mean shift procedure, a robust clustering technique, which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. A non-parametric technique for the recovery of significant image features is presented and segmentation module is developed using the mean shift algorithm to segment each image. In these algorithms, the only user set parameter is the resolution of the analysis and either gray level or color images are accepted as inputs. Extensive experimental results illustrate excellent performance.

Keywords**—**Segmentation, Clustering, Image Retrieval, Features.

I. INTRODUCTION

CBIR for genera Mean Shift procedure l-purpose image databases is a highly challenging problem. An image in a database, or a portion of an image, is represented by a set of regions, corresponding to objects, which are characterized by color, texture, shape, and location. Region-based image retrieval (RBIR) systems attempt to overcome the drawback of global features by representing images at object-level, which is intended to be close to the perception of human visual system [3], [10], [12]. We have created a database of 5000 natural images. The procedure discussed in the section 3 is applied on each image and is segmented to get the region of interest.

Low-level computer vision tasks are misleadingly difficult. Incorrect results can be easily obtained since the employed techniques often rely upon the user correctly guessing the values for the tuning parameters. To improve performance, the execution of low-level tasks should be task driven, i.e., supported by independent high-level information. This approach, however, requires that, first; the low-level stage provides a reliable extraction process be controlled only by very few tuning parameters corresponding to intuitive measures in the input domain [7].

Feature space-based analysis of images is a paradigm, which can achieve the above-stated goals. A feature space is a mapping of the input obtained through the processing of the data in small subsets at a time. For each subset, a parametric representation of the feature of interest is obtained and the result is mapped into a point in the multidimensional space of the parameter. After the entire input is processed, significant features correspond to denser regions in the feature space, i.e., to clusters, and the goal of the analysis is the delineation of these clusters [13].

Analysis of the feature space is application independent. While there are a plethora of published clustering techniques, most of them are not adequate to analyze the feature spaces derived from the real data. Methods that rely upon a priori knowledge of the number of clusters present, as well as methods which implicitly assume the same shape for all the clusters in the space, are not able to handle the complexity of a real feature space. For a survey of such methods, see [20].

Most often different variants of k-means clustering are employed, in which the feature space is represented as a mixture of normal distributions. The user usually sets the number of mixture components. The popularity of the k-means algorithm is due to its low computational complexity of $O(nkNd)$, where n is the number of points, d the dimension of the space, and N the number of iterations, which is always small relative to n. However, since it imposes a rigid delineation over the number of clusters present, the k-means clustering can return erroneous results when the embedded assumptions are not satisfied. Moreover, the k-means algorithm is not robust, points, which do not belong to any of the K clusters, can move the estimated means away from the densest regions [5], [6].

Arbitrarily structured feature spaces can be analyzed only by nonparametric methods since these methods do not have embedded assumptions. Numerous nonparametric clustering methods were described in the literature and they can be classified into two large classes: hierarchical and density estimation. Hierarchical clustering techniques either aggregate or divide the data based on some proximity measure. The

Manuscript received January 10, 2007.

M. V. Sudhamani is with the Siddaganga Institute of Technology, Tumkur, Karnataka, India, Pin: 572 103 (corresponding author to provide phone: 91-0816-2214064; fax: 91-0816-2282994; e-mail: mvsudha_raj@ hotmail.com).

C. R. Venugopal, is with Sri Jayachamarajendra college of Engineering, Mysore, Karnataka, India, Pin: 570 006 (e-mail: crv@sjce.ac.in).

hierarchical methods tend to be computationally expensive and the definition of a meaningful stopping criterion for the fusion (or division) of the data is not straightforward [2].

The rationale behind the density estimation-based non-parametric clustering approach is that the feature space can be regarded as the empirical probability density function (p.d.f) of the represented parameter. Dense regions in the feature space thus correspond to local maxima of the p.d.f., that is, to the modes of the unknown density. Once the location of a mode determined, the cluster associated with it is delineated based on the local structure of the feature space [7].

The Mean shift procedure is a robust clustering technique, is also an iterative technique, but instead of the means, it estimates the modes of the multivariate distribution underlying the feature space. The number of clusters is obtained automatically by finding the centers of the densest regions in the space (the modes) [7].

Our approach to mode detection and clustering is based on the mean shift procedure. Mean shift is a simple, nonparametric clustering technique for estimation of the density gradient was proposed in 1975 by Fukunaga and Hostetler. The idea was generalized by [1]. The advantages of employing a mean shift type procedure in density estimation were only recently rediscovered. As will be proven in the sequel, a computational module based on the mean shift procedure is an extremely versatile tool for feature space analysis and can provide reliable solutions for many vision tasks.

The region representation used by the mean shift segmentation is similar to the blob representation employed in [17]. However, while the blob has a parametric description, the partition generated by the mean shift is characterized by a nonparametric model. An image region is defined by all the pixels associated with the same mode in the joint domain. In [19], a nonparametric clustering method is described in which, after kernel density estimation with a small bandwidth, the clusters are delineated through concatenation of the detected mode's neighborhoods. The merging process is based on two intuitive measures capturing the variations in the local density. Being a hierarchical clustering technique, the method is computationally expensive; it takes several minutes in MATLAB to analyze a 2,000-pixel subsample of the feature space. The method is not recommended to be used in the joint domain since the measures employed in the merging process become ineffective. Comparing the results for arbitrarily shaped synthetic data [19] with the mean shift method here shows that the use of a hierarchical approach can be successfully avoided in the nonparametric clustering paradigm. The joint domain segmentation of the color 180 X 119 image presented in Fig. 3 is also satisfactory. Compare this result with the segmentation presented in [21] obtained by recursive thresholding. In these examples, we have noticed that regions in which a small gradient of illumination exists (like the water in Fig. 3) were delineated as a single region. Thus, the joint domain mean shift-based segmentation succeeds in overcoming the inherent limitations of methods based only on gray-level or color clustering which typically oversegment small gradient regions. Result of mean shift is also compared with that of multiscale approach in [18]. Finally, we have compared the contours of the color image with those from [16], obtained through a complex global optimization for performance. In section 2, mean shift procedure is defined and its properties are analyzed. In section 3, the procedure used as a computational module and implementation issues are discussed. Section 4, discusses the results of segmentation.

II. MEAN SHIFT PROCEDURE

The modes of a random variable y are the local maxima of its probability density function $f(y)$. However, only the empirical distribution, the data points $y_{i,}$ $i = 1,........,n$ are available. To accurately determine the locations of the modes, first a continuous estimate of the underlying density $\hat{f}(y)$ has to be defined. Later that this step can be eliminated by directly estimating the gradient of the density. To estimate the probability density in y a small neighborhood is defined around y. The neighborhood usually has a simple shape: cube, sphere or ellipsoid. Let its volume be V_y, and m_y be the number of data points inside. Then the density estimate is $\hat{f}(y) = \dfrac{m_y}{nV_y}$ which can be employed in two different ways.

First, in the nearest neighbors approach, the neighborhoods (the volumes of V_y) are scaled to keep the number of points m_y constant. A mode corresponds to a location in which the neighborhood has the smallest volume [14].

Second, in the kernel density approach, the neighborhoods have the same volume V_y and the number of points m_y inside is counted. A mode corresponds to location in which the neighborhood contains the largest number of points [14].

For goal of finding the local maxima of $\hat{f}(y)$, the kernel density methods are more suitable. Kernel density estimation is a widely used technique in statistics and pattern recognition, where it is also called the Parzen window method [7].

Kernel density estimation is the most popular density estimation method. Given n data points x_i, i=1,….n in the d-dimensional space R^d, the multivariate kernel density estimator with kernel k(x) and a symmetric positive definite d X d bandwidth matrix H, computed in the point x is given by

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} k_H(x - x_i) \tag{1}$$

Where $k_H(x) = |H|^{-\frac{1}{2}}\ k(H^{-\frac{1}{2}}x)$ (2)

The d-variate kernel $k(x)$ is a bounded function with compact support satisfying

$$\int_{R^d} k(x)dx = 1 \qquad \lim_{\|x\|\to\infty} \|x\|^d\ k(x) = 0$$

$$\int_{R^d} xk(x)dx = 0 \qquad \int_{R^d} xx^T k(x)dx = c_k I \tag{3}$$

Where c_k is a constant. The multivariate kernel can be generated from a symmetric univariate kernel $k_1(x)$ in two different ways

$$k^p(x) = \prod_{i=1}^{d} k_1(x_i) \qquad k^s(x) = a_{k,d} k_1(\|x\|) \tag{4}$$

Where $k^p(x)$ is obtained from the product of the univariate kernels and $k^s(x)$ from rotating $k_1(x)\, in\, R^d$, i.e., $k^s(x)$ is radially symmetric. The constant $a_{k,d}^{-1} = \int_{R^d} k_1(\|x\|)dx$ assures that $k^s(x)$ integrates to one. Either type of multivariate kernel obeys (3), but for our purposes, the radially symmetric kernels are often more suitable. We are interested only in a special class of radially symmetric kernels satisfying

$$k(x) = c_{k,d} k\left(\|x\|^2\right) \tag{5}$$

in which case it suffices to define the function k(x) called the profile of the kernel, only for $x \geq 0$. The normalization constant $c_{k,d}$, which makes k(x) integrate to one, is assumed strictly positive. Using a fully parameterized H increases the complexity of the estimation and, in practice, the bandwidth matrix H is chosen either as diagonal H=diag $[h_1^2, \ldots, h_d^2]$, or proportional to the identity matrix $H=h^2I$. The clear advantage of the latter case is that only one bandwidth parameter h>0 must be provided. However, as can be seen from (2), then the validity of an Euclidian metric for the feature space should be confirmed first.

Employing only one bandwidth parameter, the kernel density estimator (1) becomes the well-known expression

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right) \tag{6}$$

The quality of the kernel density estimator is measured by the mean of the square error between the density and its estimate, integrated over the domain of definition. In practice, however, only an asymptotic approximation of this measure (denoted as AMISE) can be computed. Under the asymptotics, the number of data points $n \to \infty$, while the bandwidth $h \to 0$ at a rate slower than n^{-1}.For the both types of multivariate kernels, the AMISE measure is minimized by the Epanechnikov kernel having the profile

$$k_E(x) = \begin{cases} 1-x & 0 \leq x \leq 1 \\ 0 & x > 1, \end{cases} \tag{7}$$

which yields the radially symmetric kernel.

$$k_E(x) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2)(1-\|x\|^2) & \|x\| \leq 1 \\ 0 & otherwise, \end{cases} \tag{8}$$

Where c_d is the volume of the unit d- dimensional sphere. Note that the Epanechnikov profile is not differentiable at the boundary. The profile

$$k_N(x) = \exp(-\frac{1}{2}x) \qquad x \geq 0 \tag{9}$$

Yields the multivariate normal kernel.

$$k_N(X) = (2\pi)^{-\frac{d}{2}} \ \exp(-\frac{1}{2}\|x\|^2) \tag{10}$$

for both types of composition (4). The normal kernel is often symmetrically truncated to have a kernel with finite support. While these two kernels will suffice for most applications we are interested in, all the results presented below are valid for arbitrary kernel within the conditions to be stated. Employing the profile notation, the density estimator (6) can be rewritten as

$$\hat{f}_{h,k}(x) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^{n} k\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \tag{11}$$

The first step in the analysis of the feature space with the underlying density $f(x)$ is to find the modes of this density. The modes are located along the zeros of the gradient $\nabla f(x) = 0$ and the mean shift procedure is an elegant way to locate these zeros without estimating the density.

A. *Density Gradient Estimation*

The density gradient estimator is obtained as the gradient of the density estimator by exploiting the linearity of (11)

$$\hat{\nabla} f_{h,k}(x) \equiv \nabla\hat{f}_{h,k}(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (x - x_i) k'\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \tag{12}$$

We define the function

$$g(x) = -k'(x), \tag{13}$$

Assuming that the derivative of the kernel profile k exists for all $x \in [0, \infty)$, except for a finite set of points. Now, using $g(x)$ for profile, the kernel $G(x)$ is defined as

$$G(x) = c_{g,d}\, g\left(\|x\|^2\right), \tag{14}$$

Where $c_{g,d}$ is the corresponding normalization constant. The kernel $k(x)$ was called the shadow of $G(x)$ in a slightly different context. Note that Epanechnikov kernel is the shadow of the uniform kernel, i.e., the d-dimensional unit sphere, while the normal kernel and its shadow have the same expression. Introducing $g(x)$ into (12) yields,

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

$$\hat{\nabla} f_{h,k}(x) = \frac{2c_{k,d}}{nh^{d+2}} \left[\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)\right] \left[\frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x\right] \tag{15}$$

Where $\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)$ is assumed to be a positive number. This condition is easy to satisfy for all the profiles met in practice. Both terms of the product in (15) have special significance. From (11), the first term is proportional to the density estimate at x computed with the kernel G.

$$\hat{f}_{h,G}(x) = \frac{c_{g,d}}{nh^d} \sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \tag{16}$$

The second term is the mean shift.

$$m_{h,G}(x) = \frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \quad , \tag{17}$$

i.e., the difference between the weighted mean, using the kernel G for weights, and x, the center of the kernel (window). From (16) and (17),(15) becomes

$$\hat{\nabla} f_{h,k}(x) = \hat{f}_{h,G}(x)\frac{2c_{k,d}}{h^2 c_{g,d}} m_{h,G}(x) \tag{18}$$

yielding

$$m_{h,G}(x) = \tfrac{1}{2} h^2 c \frac{\hat{\nabla} f_{h,k}(x)}{\hat{f}_{h,G}(x)} \tag{19}$$

The expression (19) shows that, at location x, the mean shift vector computed with kernel G is proportional to the normalized density gradient estimate obtained with the kernel K. The normalization is by the density estimate in x is computed with the kernel G. The mean shift vector thus always points toward the direction of maximum increase in the density. This is a more general formulation of the property first remarked by Fukunaga and Hostetler.

The relation captured in (19) is intuitive; the local mean is shifted towards the region in which the majority of the points reside. Since the mean shift vector is aligned with the local gradient estimate, it can define a path leading to a stationary point of the estimated density. The modes of the density are such stationary points. The mean shift procedure, obtained by successive

- Computation of the mean shift vector $m_{h,G}(x)$,
- Translation of the kernel (window) $G(x)$ by $m_{h,G}(x)$,

is guaranteed to converge at a nearby point where the estimate(11) as zero gradient. The presence of normalization by the density estimate is a desirable feature. The regions of the low-density values are of no interest for the feature space analysis and, in such regions, the mean shift steps are large. Similarly, near local maxima the steps are small and the analysis more refined. The mean shift procedure thus is an adaptive gradient ascent method [4], [8].

B. *Sufficient Condition for Convergence*

Denote by $\{y_j\}_{j=1,2..}$ the sequence of successive locations of the kernel G, where, from (17),

$$y_{j+1} = \frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} \qquad j = 1,2,..... \tag{20}$$

is the weighted mean at y_j computed with kernel G and y_1 is the center of the initial position of the kernel. The corresponding sequence of density estimates computed with kernel k, $\left\{\hat{f}_{h,k}(j)\right\}_{j=1,2,.......,}$, is given by

$$\hat{f}_{h,k}(j) = \hat{f}_{h,k}(y_j) \qquad j = 1,2,..... \tag{21}$$

A kernel k that obeys some mild conditions suffices for the convergence of the sequences $\{y_j\}_{j=1,2,..}$ and $\left\{\hat{f}_{h,k}(j)\right\}_{j=1,2,....,}$.

C. *Mean Shift-Based Mode Detection*

Let us denote by y_c and $\hat{f}^{c}_{h,k} = \hat{f}_{h,k}(y_c)$ the convergence points of the sequences $\{y_j\}_{j=1,2,..}$ and $\left\{\hat{f}_{h,k}(j)\right\}_{j=1,2,.......,}$, respectively. The implications are the following. First, the magnitude of the mean shift vector converges to zero. Indeed, from (17) and (20) the j^th^ mean shift vector is

$$m_{h,G}(y_j) = y_{j+1} - y_j \tag{22}$$

and, at the limit, $m_{h,G}(y_c) = y_c - y_c = 0$. In other words, the gradient of the density estimate (11) y_c is zero

$$\nabla \hat{f}_{h,k}(y_c) = 0, \tag{23}$$

due to (19). Hence, y_c is a stationary point of $\hat{f}_{h,k}$. Second, since $\left\{\hat{f}_{h,k}(j)\right\}_{j=1,2,.......,}$ is monotonically increasing, the mean shift iterations satisfy the conditions required by the capture theorem, which states that the trajectories of such gradient methods are attracted by local maxima if they are unique (within a small neighborhood) stationary points. That is, once y_j gets sufficiently close to a mode of $\hat{f}_{h,k}$, it converges to it. The set of all locations that converge to the same mode defines the basin of attraction of that mode.

The theoretical observations from above suggest a practical algorithm for mode detection:

- Run the mean shift procedure to find the stationary points of $\hat{f}_{h,k}$,
- Prune these points by retaining only the local maxima.

The local maxima points are defined, according to the Capture theorem, as unique stationary points within some small open sphere. Perturbing each stationary point by a random vector of small norm and letting the mean shift procedure converge again can test this property. Should the point of convergence be unchanged (up to a tolerance), the point is a local maximum.

D. Smooth Trajectory Property

The mean shift procedure employing a normal kernel has an interesting property. Its path toward the mode follows a smooth trajectory, the angle between consecutive mean shift vectors being always less than 90 degrees. Using the normal kernel (10), the j th mean shift vector is given by

$$m_{h,N}(y_j) = y_{j+1} - y_j = \frac{\sum_{i=1}^{n} x_i \exp\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} \exp\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - y_j \qquad (24)$$

III. COLOR IMAGE SEGMENTATION

The versatility of the feature space analysis enables the design of algorithms in which the user controls performance through a single parameter, the resolution of the analysis (i.e., bandwidth of the kernel). Both gray level and color images are processed with the same algorithm, in the former case, the feature space containing two degenerate dimensions that have no effect on the mean shift procedure. We have chosen the $L^*u^*v^*$ space, whose coordinates are related to the RGB values by nonlinear transformations [15]. The daylight standard D_{65} was used as reference illuminant.

A. Mean Shift Filtering

Let x_i and $z_i, i = 1,....,n,$ be the d-dimensional input and filtered image pixels in the joint spatial-range domain. For each pixel,

1. Initialize $j = 1$ and $y_{i,1} = x_i$.
2. Compute $y_{i,j+1}$ according to (20) until convergence, $y = y_{i,c}$.
3. Assign $z_i = (x_i^s, y_{i,c}^r)$.

The superscripts s and r denote the spatial and range components of a vector, respectively. The assignment specifies that the filtered data at the spatial location x_i^s will have the range component of the point of convergence $y_{i,c}^r$.

The kernel in the mean shift procedure moves in the direction of the maximum increase in the joint density gradient, while the bilateral filtering uses the fixed, static window. In the image smoothed by mean shift filtering, information beyond the individual windows is also taken into account.

B. Image Segmentation with Mean Shift

Large classes of image segmentation algorithms are based on feature space analysis. In this paradigm the pixels are mapped into a color space and clustered, with each cluster delineating a homogeneous region in the image i.e., a piecewise constant model is enforced over the image.

When the mean shift procedure is applied to every point in the feature space the points of convergence aggregate in groups, which can be merged. These are the detected modes, and the associated data points define their basin of attraction. The clusters are delineated by the boundaries of the basins, and thus can have arbitrary shapes. The number of significant modes detected automatically determines the number of significant clusters present in the feature space.

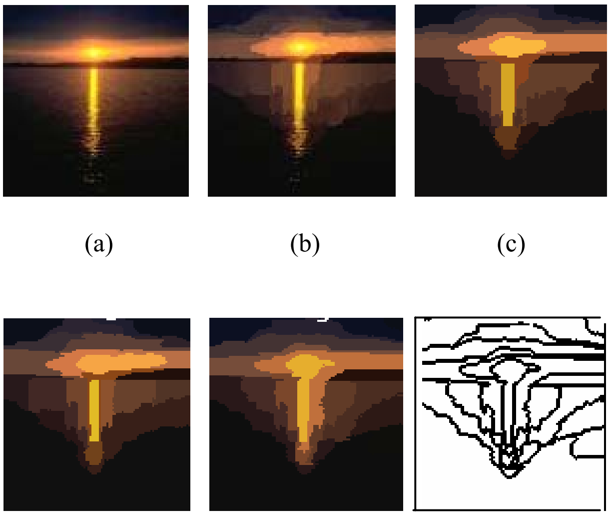
The color image segmentation algorithm proposed in [8], a five-dimensional feature space was used. The $L^*u^*v^*$ color space was employed since its metric is a satisfactory approximation to Euclidean, thus allowing the use of spherical windows. The remaining two dimensions were the lattice coordinates. A cluster in this 5D feature space thus contains pixels that are not only similar in color but also contiguous in the image.

Recursive application of the mean shift property yields a simple mode detection procedure. The modes are the local maxima of the density, i.e. $\nabla f(x) = 0$. They can be found by moving at each iteration window by the mean shift vector, until the magnitude of the shifts becomes less than a threshold. The procedure is guaranteed to converge.

The quality of segmentation is controlled by the spatial h_s, and the color h_r, resolution parameters defining the radii of the (3D/2D) windows in the respective domains. The segmentation algorithm has two major steps. First, the image is *filtered* using mean shift in 5D, replacing the value of each pixel with the 3D (color) component of the 5D mode it is associated to. Note that the filtering is discontinuity preserving. In the second step, the basins of attraction of the modes, located within $h_r/2$ in the color space are recursively fused until convergence. The resulting large basins of attraction are the delineated regions, and the value of all the pixels within is set to their average. It is important to emphasize that the segmenter processes gray level and color images in the same way. The only difference is that in the former case the feature space has three dimensions, the gray value and the lattice coordinates. To optimize the performance, in the filtering step, a speed-up of about four to five times relative to the actual time is obtained by not applying the mean shift procedure to the pixels which are on the mean shift trajectory of another (already processed) pixel. These pixels were directly associated with the mode to which the path converged. The approximation does not yield a visible change in the filtered image.

In the fusion step, extensive use was made of region adjacency graphs and graph contraction with a union- find algorithm [11]. The initial region adjacency graph was built from the filtered image, the modes being the vertices of the graph and the edges were defined based on four connectivity on the lattice. The fusion was performed as a transitive closure operation on the graph, under the condition that the color difference between two adjacent nodes should not exceed h_r/2. At convergence, the color of the regions was recomputed and the transitive closure was again performed. After at most three iterations the final labeling of the image (segmentation) was obtained. Small regions were then allocated to the nearest neighbor in the color space. Note that employing a look-up table, which captures the relation between the smallest significant color difference and the minimum region size. The implementation of the color image segmenter was also tested for equivariance under 90° rotations on the lattice. That is, when the input image is rotated the segmented image rotates accordingly. This property assures that the output of the processing does not depend on the order in which the pixels in the image are processed.

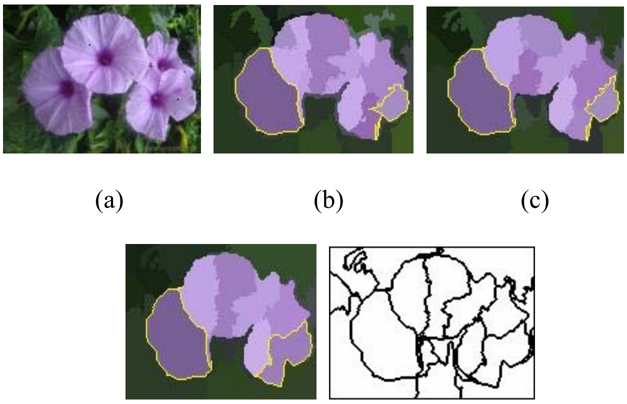
IV. RESULTS


(a) (b) (c)

(d) (e) (f)

Fig. 1 Mean Shift segmentation. (a) Input color Image of 99 X 132 (b) Filtered Image. (c) Segmentation at higher resolution i.e. h_s=7, h_r=6.5 , M =100. (d) Segmentation at lower resolution with h_s=7, h_r=7.5 M =100. (e)(f) Segmentation with h_s=9, h_r=6.5, M =200 and region boundaries

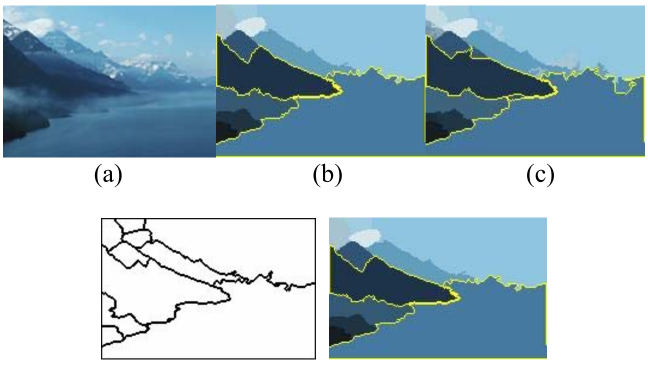
The performance of the mean shift segmentation using lattice is accessed with seven images of different natures. The 99 X 132 color scenery image in Fig. 1a is segmented by setting different values to the input parameters h_s, h_r and M . When the image is segmented at a lower resolution i.e. h_s=7, h_r=7.5 M =100 only a very few features are recovered i.e. obtained 21 regions as in Fig. 1d. On the other hand, when the resolution is slightly increased i.e. h_s=7, h_r=6.5 M =100 a significant clutter appears i.e. obtained 24 regions as in Fig. 1c. M , Minimum Number of pixels in the region also changes the recovery of features as in Fig. 1e and 1f. Here h_s=9, h_r=6.5 M =200 and obtained regions are 16. Tested also for h_s=7, h_r=6.5, M =220 obtained regions are 13. Time taken for segmentation is 0.59 secs.


(a) (b) (c)

(d) (e)

Fig. 2 Mean Shift segmentation. (a) Input color Image of 150 X 112 (b) Segmentation at higher resolution i.e. h_s=7, h_r=6.5, M =200. (c) Segmentation at lower resolution with h_s=7, h_r=7.5 M =250. (d) (e) Segmentation at lower resolution with h_s=8, h_r=7.5 M =350 and Region boundaries

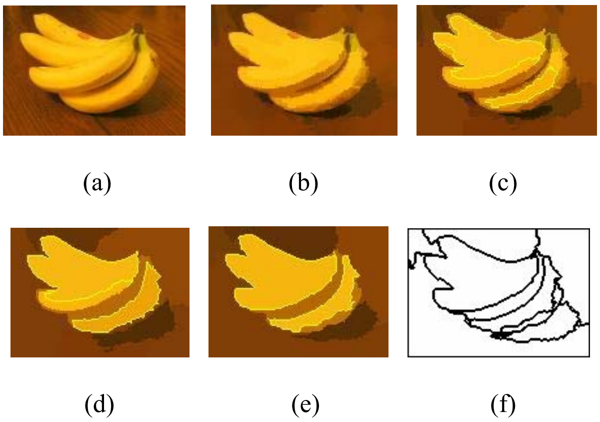
The 150 X 112 color image in Fig. 2a is segmented by mean shift. When the image is segmented at a lower resolution i.e. h_s=7, h_r=7.5 M =250 only a very few features are recovered i.e. obtained 21 regions as in Fig. 2c. Two regions of interest (region number 11 and 16) are marked and their mode values, number of pixels in the region are accessed. Time complexity of segmentation is 0.73 seconds. On the other hand, when the resolution is slightly increased i.e. h_s=7, h_r=6.5 M =200 a significant clutter appears i.e. obtained 30 regions as in Fig. 2b. Region number 18 and 23 are marked. Time complexity of segmentation is 0.71 seconds. Minimum Number of pixels in the region also changes the recovery of features as in Fig. 2d and 2e. Here h_s=8, h_r=7.5 M =350 and regions obtained are 16. The time taken to segment this image is 0.90 seconds.


(a) (b) (c)

(d) (e)

Fig. 3 Mean Shift segmentation. (a) Input color Image of 180 X 119 (b) Segmentation at higher resolution i.e. $h_s=7, h_r=6.5$, $M=20$. (c) (d) Segmentation at lower resolution with $h_s=7, h_r=7.5, M=200$. (e) Segmentation with $h_s=8, h_r=7.5$, $M=200$

Fig. 3 is a 180 X 119 natural scenery image 3a is segmented with $h_s=7, h_r=6.5$ $M=20$ and obtained 23 regions as in 3b. Time taken for segmentation is 0.98 seconds. Regions of interest are marked and their corresponding color features are extracted. 3c and 3d are the results with $h_s=7, h_r=7.5$, $M=200$ with time complexity of 1.0 second and obtained 10 regions. Tested also for $h_s=8, h_r=7.5, M=200$ and obtained 11 regions. Time taken for segmentation is 1.0 seconds as in Fig. 3e.


(a) (b) (c)

(d) (e) (f)

Fig. 4 Mean Shift segmentation. (a) Input color Image of 140 X 100 (b) Filtered Image. (c) Segmentation at higher resolution i.e. $h_s=7, h_r=6.5$, $M=20$. (d) Segmentation at lower resolution with $h_s=7, h_r=6.5$, $M=300$. (e) (f) Segmentation with $h_s=7, h_r=8.5$, $M=250$

Fig. 4 is a 140 X 100 color image. This is tested with $h_s=7, h_r=6.5$ $M=20$ and obtained 75 regions as in 4c. We have accessed number of pixels and color features of 13th and 41st region with time complexity of 0.65 seconds. In fig. 4d the value of M is increased to 300 and obtained 10 regions with 0.65 seconds. Fig. 4e and 4f are segmented with $h_s=7, h_r=8.5, M=250$ and obtained 9 regions. Time taken is 0.68 seconds.

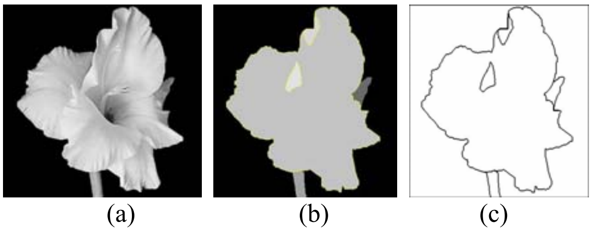

(a) (b) (c)

Fig. 5 Mean Shift segmentation. (a) Input gray level Image of 400 X 390 (b) Segmentation at lower resolution i.e. $h_s=9, h_r=7.5$, $M=600$. (c) Region boundaries

Fig. 5 is a 400 X 390 gray level image. This is tested with $h_s=9, h_r=7.5$ $M=600$ and obtained 6 regions as in 5b and 5c shows the corresponding region boundaries. Time taken is 9.18 seconds. Tested also for $h_s=7, h_r=6.5$ $M=20$ and number regions obtained are 53 with 6.7 secs. With high speed time taken is 0.64 secs for same parameters. By changing parameters values to $h_s=7, h_r=7.5$ $M=200$ with high speed obtained regions are 22 with 0.57secs.

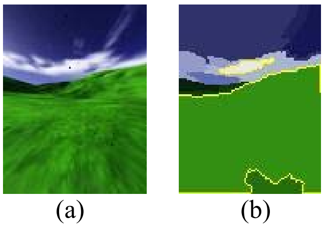

(a) (b)

Fig. 6 Mean Shift segmentation. (a) Input color nature image of 89 X 118. (b) Segmentation at lower resolution i.e. $h_s=7, h_r=9.5$, $M=100$. Obtained 11 regions with 0.57 secs

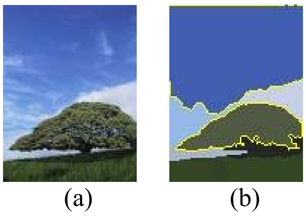

(a) (b)

Fig. 7 Mean Shift segmentation. (a) Input color nature image of 89 X 118 (b) Segmentation at lower resolution i.e. $h_s=7, h_r=9.5$, $M=100$

Fig. 6 and Fig. 7 also show the excellent performance of the mean shift algorithm. The Fig. 7 is segmented with h_s=7, h_r=9.5, M =100. Time taken is 1.96 secs. With high speed time taken is 0.49 secs.

We can observe from the above results that the values of the parameters significantly change the performance of segmentation. So the parameters values must be chosen carefully to retain the quality of segmentation. We have accessed number of pixels in interested cluster (region) to have percentage, color feature (mode value) associated with each cluster, how many such clusters forms a required region/object ect., to store them in the database for image retrieval.

V. CONCLUSION

The simplicity of the basic computational module, the mean shift algorithm, enables the feature space analysis to be accomplished very fast. The segmentation is not very sensitive to choice of h_s and h_r. The range parameter h_r and the smallest significant feature size M control the number of regions in the segmented image. An important advantage of mean shift-based segmentation is its modularity that makes the control of segmentation output very simple. The more an image deviates from the assumed piecewise constant model, larger values have to be used for h_r and M to discard the effect of small local variations in the feature space. The procedure is implemented to get the required information for content-based image retrieval. The results of the segmentation are provided and discussed for different images in the previous section. Relevant feature like color (mode value), its percentage, pixels belongs to a cluster (Region / boundary information) are extracted for the interested regions after segmentation and intern stored in the database.

Kernel density estimation, in particular, and nonparametric techniques, in general, does not scale well with the dimension of the space. This is mostly due to the empty space phenomenon by which most of the mass in a high-dimensional space is concentrated in a small region of the space. Thus, whenever the feature space has more than (say) six dimensions, the analysis should be approached carefully.

To conclude, the mean shift procedure is a valuable computational technique whose versatility can make it an important component in computer vision applications.

REFERENCES

[1] Y. Cheng, Mean shift, mode seeking, and clustering, *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 17, 790-799, 1995.

[2] J.-M. Jolion, P. Meer, S. Bataouche, Robust clustering with applications in computer vision, IEEE Trans. Pattern Anal. Machine Intell vol. 13, 791-802, 1991.

[3] W. Skarbek, A. Koschan, *Colour Image Segmentation: A Survey,* Technical Report, Technical University Berlin, October 1994.

[4] Dorin Comaniciu Peter Meer, Robust Analysis of Feature Spaces: Color Image Segmentation, *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* San Juan, Puerto Rico, pp. 750-755, June 1997.

[5] Arnaldo J. Abrantes and Jorge S. Marques, The Mean Shift Algorithm and the Unifie Framework , *Proceedings of the 17th International Conference on Pattern Recognition* –2004 (ICPR'04).

[6] B. Georgescu, I. Shimshoni and P. Meer, Mean shift based clustering in high dimensions: A Texture classification example, *Proc. Ninth IEEE International Conference on Computer Vision*, pp. 456–463, Oct. 2003.

[7] Peter Meer, Gerard Medioni and Sing Bing Kang, *Robust techniques for computer vision (*Prentice Hall, 2004).

[8] Dorin Comaniciu and Peter Meer, Mean Shift Analysis and Applications,*7th Int'l Conf. on Comp. Vis.,* Kerkyra, Greece, 1197-1203, Sep. 1999.

[9] Dorin Comaniciu and Visvanathan Ramesh ,Real-Time Tracking of Non-Rigid Objects using Mean Shift, *IEEE CVPR,* 2000.

[10] Jeff Strickrott, *A Survey of Image Segmentation Techniques for content-based retrieval of multimedia data,* Department of Computer Science, Florida International University.

[11] R. Sedgewick. *Algorithms in C.* Addison-Wesley, pp.441-449, 1990.

[12] James W.wang, *Integrated Region-Based Image Retrieval,* Kluwer academic publishers, 2001

[13] Richard O.Duda, peter E. Hart, David G. stock, *Pattern classification,* wiley, 2002,

[14] Forsyth and Ponce, *A Computer Vision. A modren Approach,* Prentice Hall, 2003.

[15] Werner Bailera, Peter Schallauera, Harald Bergur Haraldssonb, Herwig Rehatscheka, Optimized Mean Shift Algorithm for Color Segmentation in Image Sequences, *Proc. Conference on Image and Vid Communications and Processing,* IS&T/SPIE Electronic Imaging, San Jose, CA, USA, Jan. 2005.

[16] S.C Zhu and A.Yuille , Region competition: Unifying Snakes, Region Growing, and Bayes/MDL for multiband Image Segmentation, *IEEE Trans. Pattern analysis and Machine Intelligence, Vol. 18*, no.9, pp.884-900, Sept. 1996.

[17] C. Wren, Azarbayejani, T. Darrell, and A. Pentland, pfinder: Real_Time Tracking of the Human Body, *IEEE trans. Pattern Analysis and Machine Intelligence*, Vol. 19, no.7, pp.780-785, July 1997.

[18] M. Tabb and N. Ahuja, Multiscale Image Segmentation by Integrated Edge and region Detection, *IEEE Trans. Image Processing*, vol. 6, pp.642-655, 1997.

[19] E.J. Pauwels and G.Frederix., Finding Salient Regions in Images, *Computer vision and Image Understanding* , vol. 75, pp. 73-85,1999.

[20] A.K .Jain , R.P.W. Duin, and J.Mao, Statistical Pattern Recognition: A Review, *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol.22, no.1, pp. 4-37, Jan 2000.

[21] Y. Ohta, T.Kanade, and T.Sakai, Color Information for Region Segmentation*, Compute Graphics and Image Processing, vol.13,* pp.222-241, 1980.

M. V. Sudhamani is a doctoral student at Visvesvaraya Technological University, Karnataka, India. She has received her M.Tech degree in Computer Science & Eng. from Mangalore University and B.E degree in Computer Science & Eng. from Bangalore University. She is currently working as an Assistant Professor in Information Science & Engineering Department., at Siddaganga Institute of Technology, Tumkur. Her main research interests include Content- based image retrieval, Image processing, Multimedia computing and computer vision.

Dr. C. R. Venugopal is currently working as Professor in Electronics & Communication Department, S.J. College of Engineering, Mysore. He has obtained his M.E Degree from Indian Institute of Science, Bangalore. He has received his PhD from Indian Institute of Technology, Bombay, India. His research interests include Multimedia Computing and Communication, Information Retrieval, Computer vision, Pattern Recognition and Security in Digital Storage Area Network. His other activities are IEEE Staff Advisor, IETE Journal Reviewer, and IEEE Software Journal Reviewer.