Department of Electrical
and
Computer Systems Engineering

Technical Report
MECSE-1-2003

False-Peaks-Avoiding Mean Shift Method for Unsupervised Peak-ValleySliding Image Segmentation

H. Wang and D. Suter

MONASH
UNIVERSITY

False-Peaks-Avoiding Mean Shift Method for Unsupervised Peak-Valley Sliding Image Segmentation

Hanzi Wang and David Suter, IEEE, Senior Member
Department of Electrical and Computer Systems Engneering
Monash University, Clayton Vic. 3800, Australia.
{hanzi.wang ; d.suter}@eng.monash.edu.au

Abstract

The mean shift (MS) algorithm is sensitive to local peaks. In this paper, we show both empirically and analytically that when using sample data, the reconstructed PDF may have false peaks. We show how the occurrence of the false peaks is related to the bandwidth h of the kernel density estimator, using a one-dimensional example motivated by gray-level image segmentation. It is well known that in MS-based approaches, the choice of h is important. However, we provide a quantitative relationship between the appearance of false peaks and the value of h. For the gray-level image segmentation problem, we not only show how to avoid the false peak problem, but also we provide a complete unsupervised peak-valley sliding algorithm for gray-level image segmentation. However, the main contribution of the paper remains the characterization of the false peak problem and the questions it raises regarding this issue in more general settings (e.g. higher dimensional problems).

1. Introduction

The mean shift method has become a popular method for a wide variety of applications: video tracking [7], image filtering [6], clustering [5] and image segmentation [4][8] for example. In essence, it is a local (and thereby somewhat robust) form of mode seeking. It is local because it operates on a window and it also achieves a degree of scale selectivity since it works with a smoothed estimate of the underlying density function. In the most commonly used form [9][8], the window size and the smoothing are directly related to a quantity h that is the "bandwidth" choice for the kernel density estimator employed.

Although many authors of papers that employ the mean shift method have remarked that the value h needs to be chosen with care (and, likewise in the general literature on the underlying kernel density estimator literature), the general impression given is that the results are not that sensitive to the choice of h and that one generally takes a pragmatic "hit and miss" affair. In this paper we illustrate that there are two issues affected by the setting of h: the rather disastrous appearance of false peaks (where the application of the mean shift process will simply fail) and the choice of scale (affecting the significance of actual peaks in the underlying density – at large scales the density is very smoothed and local peaks are disregarded or merged). The latter behavior is much more benign and, indeed, as it performs a type of controlled scale-space analysis, can be used to advantage. The former is to be avoided at all costs as it will result in completely arbitrary results.

In this paper, we choose for simplicity the problem of histogram based gray level image segmentation. We show that one can rather simply predict values of h that will be problematic and thereby, in this setting, we provide a means for a completely automated approach, negating the need for the setting of any of the value of all parameters, including h (except that one may repeat the solution with a range of h to perform a type of scale space analysis).

Our novel "peak-valley sliding method" for identifying modes and anti-modes, is of interest in its own right. Moreover, the application of this method to automatically threshold multi-modal gray-level images may also be of independent interest as the method does not require the priori knowledge of the number of the peaks and valley and it is computationally effective. However, the shortcomings of gray-level image segmentation are well known and somewhat limit the direct application to situations with controlled or fortuitous lighting and reflectance content.

The challenge remaining is to analyze higher dimensional settings of the mean-shift approach to determine the false-peak behavior in such settings.

2. Density Gradient Estimation and the Mean Shift Method

There are several nonparametric methods available for probability density estimation: histogram, naive method, the nearest neighbor method, and kernel estimation [11]. The kernel estimation method is one of the most popular techniques used in estimating density. Given a set of n data points $\{x_i\}_{i=1,\ldots,n}$ in a d-dimensional Euclidian space R^d, the multivariate kernel density estimator with kernel K and window radius (band-width) h is defined as follows [11, p.76]

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

The kernel function K(x) should satisfy some conditions [13, p.95]. There are several different kinds of kernels. The Epanechnikov kernel [11, p.76] is one optimum

kernel which yields minimum mean integrated square error (MISE):

$$K_e(\mathrm{X}) = \begin{cases} \frac{1}{2} c_d^{-1}(d+2)(1-\|\mathrm{X}\|^2) & if\ \|\mathrm{X}\| < 1 \\ 0 & otherwise \end{cases} \quad (2)$$

where c_d is the volume of the unit d-dimensional sphere, e.g., $c_1=2$, $c_2=\pi$, $c_3=4\pi/3$.

The estimate of the density gradient can be defined as the gradient of the kernel density estimate (1).

$$\hat{\nabla} f(x) \equiv \nabla \hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} \nabla K(\frac{x - x_i}{h}) \quad (3)$$

According to (3), the density gradient estimate of the Epanechnikov kernel can be written as:

$$\hat{\nabla} f(x) = \frac{n_x}{n(h^d c_d)} \frac{d+2}{h^2} \left(\frac{1}{n_x} \sum_{x_i \in S_h(x)} [x_i - x] \right) \quad (4)$$

where the region $S_h(x)$ is a hypersphere of the radius h, having the volume $h^d c_d$, centered at x, and containing n_x data points.

The mean shift vector $M_h(x)$ is defined as:

$$\mathrm{M}_h(\mathrm{x}) \equiv \frac{1}{n_x} \sum_{x_i \in S_h(x)} [x_i - x] = \frac{1}{n_x} \sum_{x_i \in S_h(x)} x_i - x \quad (5)$$

Noting that $\frac{n_x}{h^d c_d}$ can be thought of as an estimate of $f(x)$, albeit using a constant valued weight over a spherical region, equation (4) can be rewritten as:

$$\mathrm{M}_h(\mathrm{x}) \equiv \frac{h^2}{d+2} \frac{\hat{\nabla} f(x)}{\hat{f}(x)} \quad (6)$$

Equation (6) firstly appeared in [10]. The equations (5) and (6) show that the mean shift vector is the difference between the local mean and the center of the window, and also that it is an estimate of the normalized density gradient: the mean shift is an unsupervised nonparametric estimator of density gradient. Applying the mean shift leads to the steepest ascent with a varying step size that is the magnitude of the gradient [3]. Since its introduction by Fukunaga and Hostetler, the mean shift method has been extensively exploited and applied in low level computer vision tasks [3][4][6][8][10][12] for its ease and efficiency.

The converged centers (or windows) correspond to modes (or centers of the regions of high concentration) of data. The proof of the convergence of the mean shift algorithm can be found in [5][8].

Almost all published methods, which employ mean shift, use its hill climbing property to find the peaks in feature spaces. However, sometimes it is very important to find the valleys in the feature spaces. For example, in histogram analysis, when modes have been recognized, valleys between modes need to be found to set thresholds to partition images [2].

2.1. Mean Shift Valley Algorithm

One characteristic of the mean shift vector is that it always points towards the direction of the maximum increase in the density. Thus the opposite direction of the mean shift vector will always points toward to a local minimum density.

In order to find valley in density space, we define the mean shift valley vector:

$$\mathrm{MV}_h(x) = -\mathrm{M}_h(x) = x - \frac{1}{n_x} \sum_{x_i \in S_h(x)} x_i \quad (7)$$

Replace $\mathrm{M}_h(x)$ in (6) by $\mathrm{MV}_h(x)$, we can obtain:

$$\mathrm{MV}_h(x) \equiv -\frac{h^2}{d+2} \frac{\hat{\nabla} f(x)}{\hat{f}(x)} \quad (8)$$

$\mathrm{MV}_h(x)$ always points towards the direction of the maximum decrease in the density.

In practice, the step-size given by the above analysis may lead to oscillation. We have observed this particularly when finding valleys (although the potential exists when seeking peaks as well). Thus we derive a recipe for avoiding the oscillations in valley seeking .

Let $\{y_k\}_{k=1,2\ldots}$ be the sequence of successive locations of the mean shift valley procedure, then we have for each k=1,2…

$$y_{k+1} = y_k + p \cdot MV_h(y_k) \quad (9)$$

p is a correction factor, and $0 < p \le 1$. If the shift step at y_k is large, it causes y_{k+1} to jump over the local valley and thus oscillate over the valley. This problem can be avoided when we adjust the correction factor p so that $\mathrm{MV}_h(y_k)^T \mathrm{MV}_h(y_{k+1})>0$.

The mean shift valley algorithm can be described as:

1. Choose the radius of the search window, set p =1, and initialize the location of the window
2. Compute the shift step vector $\mathrm{MV}_h(y_k)$.
3. Compute y_{k+1} by equation (9) and $\mathrm{MV}_h(y_{k+1})$.
4. If $\mathrm{MV}_h(y_k)^T \mathrm{MV}_h(y_{k+1})>0$, go to step 5; Otherwise, we let $p=p/2$. Repeat step 3 and 4 until $\mathrm{MV}_h(y_k)^T \mathrm{MV}_h(y_{k+1})>0$;
5. Translate the search window by $p \cdot MV_h(y_k)$.
6. Repeat step 3 to step 5 until convergence.

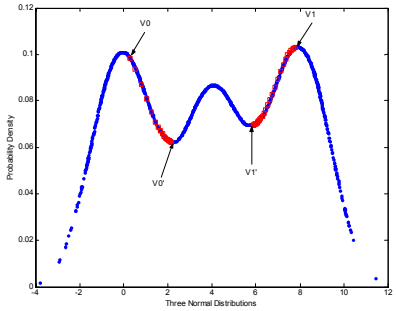


Figure 1. An example of application of the mean shift valley method to find local minimum density.

To illustrate the mean shift valley method, three normal modes (mode 1 includes 600 data points, mode 2 includes 500 data points, and mode 3 includes 600 data points) with total 1700 data points were generated in figure 1. We selected two initial points: V0 (0.3) and V1 (7.8). The search window radius was chosen as 2.0. The mean shift valley method automatically found the local minimum densities (converged points). Precisely, V0' was located at 2.1831, and V1' was at 5.8898. The centers (V0' and V1') of the converged windows correspond to the local minimum probability densities. If we use V0' and V1' as two density thresholds, the whole data can be decomposed into three modes (mode 1 corresponds to the data points whose x values are smaller than V0'; mode 2 corresponds to the data points whose x values are larger or equal than V0', but smaller than V1'; mode 3 corresponds to the data points whose x values are larger or equal than V1'). Table 1 gives the obtained parameters.

	Mode 1		Mode 2		Mode 3	
	Mean	Number	Mean	Number	Mean	Number
Generated Data	0	600	4	500	8	600
Estimated Parameters	-0.0736	603	4.0419	488	7.9592	609

Table 1. Applying the mean shift valley to decompose data.

There is one exceptional case: when there are no local valleys (e.g., uni-modal), the mean shift valley method is divergent. This can easily be avoided by terminating when no samples fall within the window.

3. The Relationship between the Gray-Level Histogram of Image and the Mean Shift.

If we are segmenting a gray-level image, the mean-shift equations can be rewritten as functions on the image intensity histogram:

$$\hat{f}(x) = \frac{d+2}{2nh^{d+2}c_d} \sum_{t_i \in S_h(x)} H(t_i)\left(h^2 - \|x - t_i\|^2\right) \qquad (10)$$

where $H(t_i)$ be the histogram of image pixels at gray level t_i (t_i is an integer and $0 \le t_i \le 255$).

The kernel density function in equation (10) is related to some discrete gray levels $\{t_i \mid t_i \in S_h(x)\}$ and the corresponding histogram $\{H(t_i) \mid t_i \in S_h(x)\}$.

Likewise:

$$\hat{\nabla} f(x) = \frac{1}{n(h^d c_d)} \frac{d+2}{h^2} \left[\sum_{t_i \in S_h(x)} H(t_i)(t_i - x) \right] = \frac{\sum_{t_i \in S_h(x)} H(t_i)}{n(h^d c_d)} \frac{d+2}{h^2} \left(\frac{\sum_{t_i \in S_h(x)} H(t_i) t_i}{\sum_{t_i \in S_h(x)} H(t_i)} - x \right) \qquad (11)$$

The last term in (11) is called the sample mean shift $M_h(x)$ in discrete gray level space:

$$M_h(x) = \frac{\sum_{t_i \in S_h(x)} H(t_i) t_i}{\sum_{t_i \in S_h(x)} H(t_i)} - x \qquad (12)$$

Equation (12) is derived from the Epanechnikov kernel. (Note: reference [14] used a Gaussian kernel - see equation 15 and 17 in that paper).

We notice also that the quantity $\frac{\sum_{t_i \in S_h(x)} H(t_i)}{n(h^d c_d)}$ is the kernel density estimate $\hat{f}(x)$ obtained with a uniform weighting in the hypersphere $S_h(x)$; thus equation (6) is consistent with equation (11) and (12).

4. The False Peak Noise.

In implementing the mean shift approach in this setting, we found, to our surprise, in some cases there are a lot of peaks appearing between two consecutive gray levels near a local maximum density (see figure 2(a) and (b)). We call these peaks the false peaks. These false peaks will seriously affect the performance of the mean shift method, i.e. the mean shift is very sensitive to these peak noises and the mean shift loop will stop at these false peaks instead of real local maximum density.

Here we analytically determine the conditions leading to this problem.

For simplicity, we have chosen a one dimensional setting. Thus d=1; c_d=2. Let $\hat{f}(t_k)$ be the kernel density estimate at gray level t_k; and let $0 < \delta x < 1$. Using equation (10) we have

$$\hat{f}(t_k + \delta x) = \frac{3}{4nh^3} \sum_{t_i \in S_h(t_k + \delta x)} H(t_i)\left(h^2 - \|t_k + \delta x - t_i\|^2\right) \qquad (13)$$

If h is an integer (h>0) and t_k+h<255, and considering t_i has to be a series of consecutive unsigned integer, we have $\{t_i \mid t_i \in S_h(t_k + \delta x)\} = \{t_i \mid t_i \in S_h(t_k)\} \cup \{t_i \mid t_i = t_k + h\}$.

The equation (13) can be rewritten as:

$$\hat{f}(t_k + \delta x) = \frac{3}{4nh^3}\left[\sum_{t_i \in S_h(t_k)} H(t_i)\left(h^2 - \|t_k + \delta x - t_i\|^2\right) + H(t_k + h)(h^2 - \|h - \delta x\|^2) \right]$$

$$= \hat{f}(t_k) + \frac{3}{4nh^3}\left[\sum_{t_i \in S_h(t_k)} H(t_i)\left(-\|\delta x\|^2 + 2\delta x \left(\frac{\sum_{t_i \in S_h(t_k)} H(t_i) t_i}{\sum_{t_i \in S_h(t_k)} H(t_i)} - t_k \right) \right) + H(t_k + h)(-\|\delta x\|^2 + 2h\delta x) \right]$$

$$= \hat{f}(t_k) + \frac{3}{4nh^3} \sum_{t_i \in S_h(t_k)} H(t_i)\left(-\|\delta x\|^2 + 2\delta x M_h(t_k) \right) + \frac{3}{4nh^3} H(t_k + h)(-\|\delta x\|^2 + 2h\delta x) \qquad (14)$$

We let:

$$A1 = \frac{3}{4nh^3} \sum_{t_i \in S_h(t_k)} H(t_i)\left(-\|\delta x\|^2 + 2\delta x M_h(t_k) \right)$$

$$A2 = \frac{3}{4nh^3} H(t_k + h)(-\|\delta x\|^2 + 2h\delta x) \qquad (15a)$$

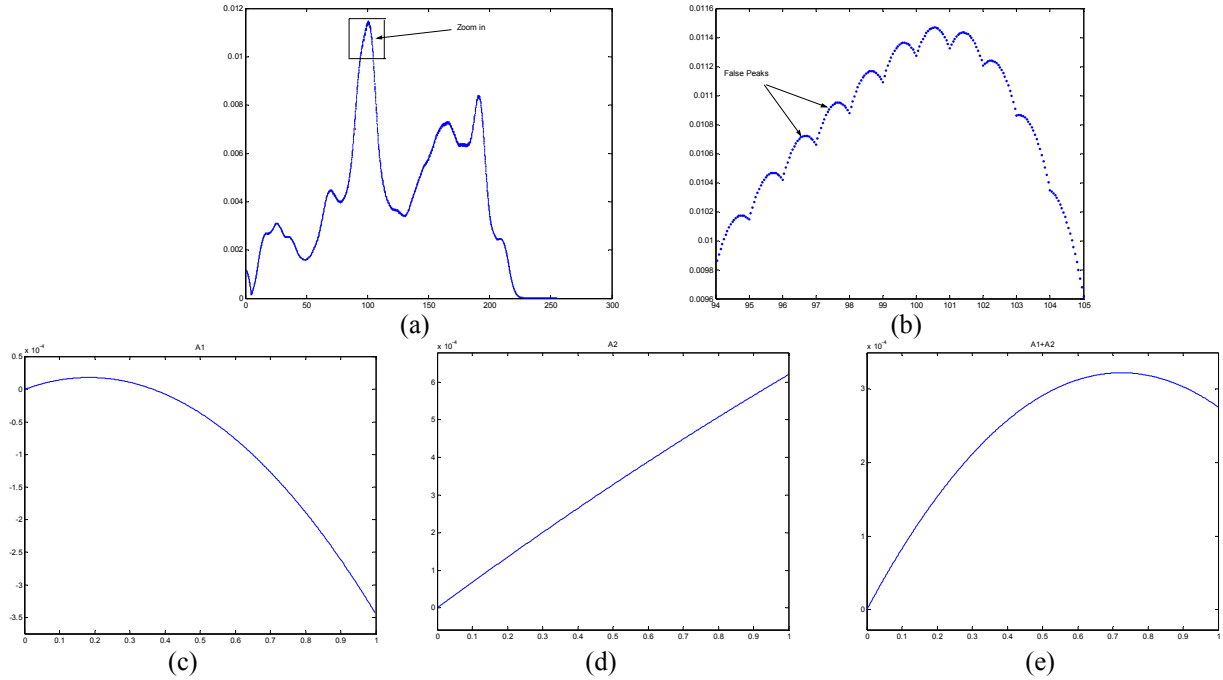


Figure 2. False peak noise. (a) Original probability density distribution with h equal to 5; (b) Zoom in a part of (a). Many false peaks introduced by A1+A2 in Eq. (15); (c)-(e) A1, A2, and A1+A2 in Eq. (15) with t_k=95;

When h>>δx, *A2* can be approximated as a linear equation (see figure 2(d)).
Equation (14) can be rewritten as:

$$\hat{f}(t_k+\delta x)=\hat{f}(t_k)+A1+A2 \qquad (15b)$$

Now we calculate the differential of $\hat{f}(t_k+\delta x)$:

$$\frac{d\hat{f}(t_k+\delta x)}{d(\delta x)}=\frac{3}{4nh^3}\sum_{t_i\in S_h(t_k)}H(t_i)(-2\delta x+2M_h(t_k))+\frac{3}{4nh^3}H(t_k+h)(-2\delta x+2h)$$

$$=\frac{3}{4nh^3}\left[-2\delta x\left(\sum_{t_i\in S_h(t_k)}H(t_i)+H(t_k+h)\right)+2\left(\sum_{t_i\in S_h(t_k)}H(t_i)M_h(t_k)+H(t_k+h)h\right)\right] \qquad (16)$$

Let (16) equal zero, we obtain:

$$\delta x=\frac{\sum_{t_i\in S_h(t_k)}H(t_i)M_h(t_k)+H(t_k+h)h}{\sum_{t_i\in S_h(t_k)}H(t_i)+H(t_k+h)} \qquad (17)$$

Substituting equation (12) into equation (17), and if $0<\delta x<1$, i.e. if:

$$\frac{\sum_{t_i\in S_h(t_k)}H(t_i)(t_k-t_i)}{H(t_k+h)}<h<\frac{\sum_{t_i\in S_h(t_k)}H(t_i)(t_k-t_i)+\sum_{t_i\in S_h(t_k)}H(t_i)+H(t_k+h)}{H(t_k+h)} \qquad (18)$$

there will be a false peak appearing between two consecutive gray level, t_k and t_k+1.
In figure 2, when we apply the mean shift to find the local maximum density with initial location at 95, we find it stopped at 95.7244, instead of the real local maximum density at 101. From (17), we obtained δx =0.7244, i.e. there is a false peak between 95 and 96.
We let L be the left item in the in equation (18) and R be the right item of (18); let $x_{MS(tk)}$ be the convergent point, obtained by the mean shift method with initial point at t_k, corresponding to the local peak. Thus if the condition: L<h<R is satisfied, we can predict that there will be a false peak between t_k and t_k+1 (see table 2).

h	L	R	δx	t_k	$x_{MS(tk)}$	False peak between t_k and t_k+1
5	-1.45	7.45	0.72	95	95.72	yes
6	-2.84	6.99	0.89	95	95.89	yes
7	-5.76	6.06	1.08	95	96.96	no
7	-5.10	7.45	0.96	96	96.96	yes
8	-9.68	3.95	1.30	95	97.94	no
8	-7.99	6.19	1.13	96	97.94	no
8	-7.42	8.96	0.94	97	97.94	yes

Table 2. False peaks prediction

The above analysis suggests that one could devise an approach that adaptively adjusts h depending upon whether false peaks are predicted. However, for this simple example, we have used a heuristic approach:
(1) for the mean shift method, we adjust the shift step as:

$$y_{k+1}=y_k+\text{ceil}(M_h(y_k)) \qquad (19a)$$

(2) for the mean shift valley method, we adjust the shift step as:

$$y_{k+1}=y_k+\text{floor}(MV_h(y_k)) \qquad (19b)$$

5 An Unsupervised Peak-Valley Sliding Algorithm for Image Segmentation.

Consider the peaks {P(i)} and valleys{V(i)}. V(0)=0 and V(n)=255. $V(0) \le P(1) < V(1) < .... < P(n) \le V(n)$.
The algorithm is described as follows:
(1) Initialise the radius h and the location of search window.
(2) Apply the mean shift method to obtain the peak P_k with the initial window location $V_{k-1}+1$.
(3) Apply the mean shift valley method to obtain valley V_k with initial window location P_k+1.
(4) Repeat step (2) and (3) until P_k or V_k is equal to or larger than 255. The questions remains as to how many of these peaks are significant. Rather than some form of scale space analysis, we post-process by:
(5) Validate peaks and valleys
(5a) Remove peaks too small compared with the largest.
(5b) Remove the smaller of two consecutive peaks if too close.
(5c) Calculate the normalized contrast [1] for a valley and two neighbouring peaks:

$$\text{Normalized Contrast} = \frac{Contrast}{Height} \qquad (20)$$

where the contrast is the difference between the smaller peak and the valley. Remove the smaller one of the two peaks if this is small.
After step 5(a)-5(c), we obtain several significant peaks {PS(1),…PS(k)}. The valleys then are chosen as the minimum of the valleys between two consecutive significant peaks. Thus we have k-1 valleys {VS(1),…VS(k-1)}.
(6) Using the obtained valleys finally obtain k segmented images by {[0, VS(1)], [VS(1), VS(2)], … [VS(k), 255]}.

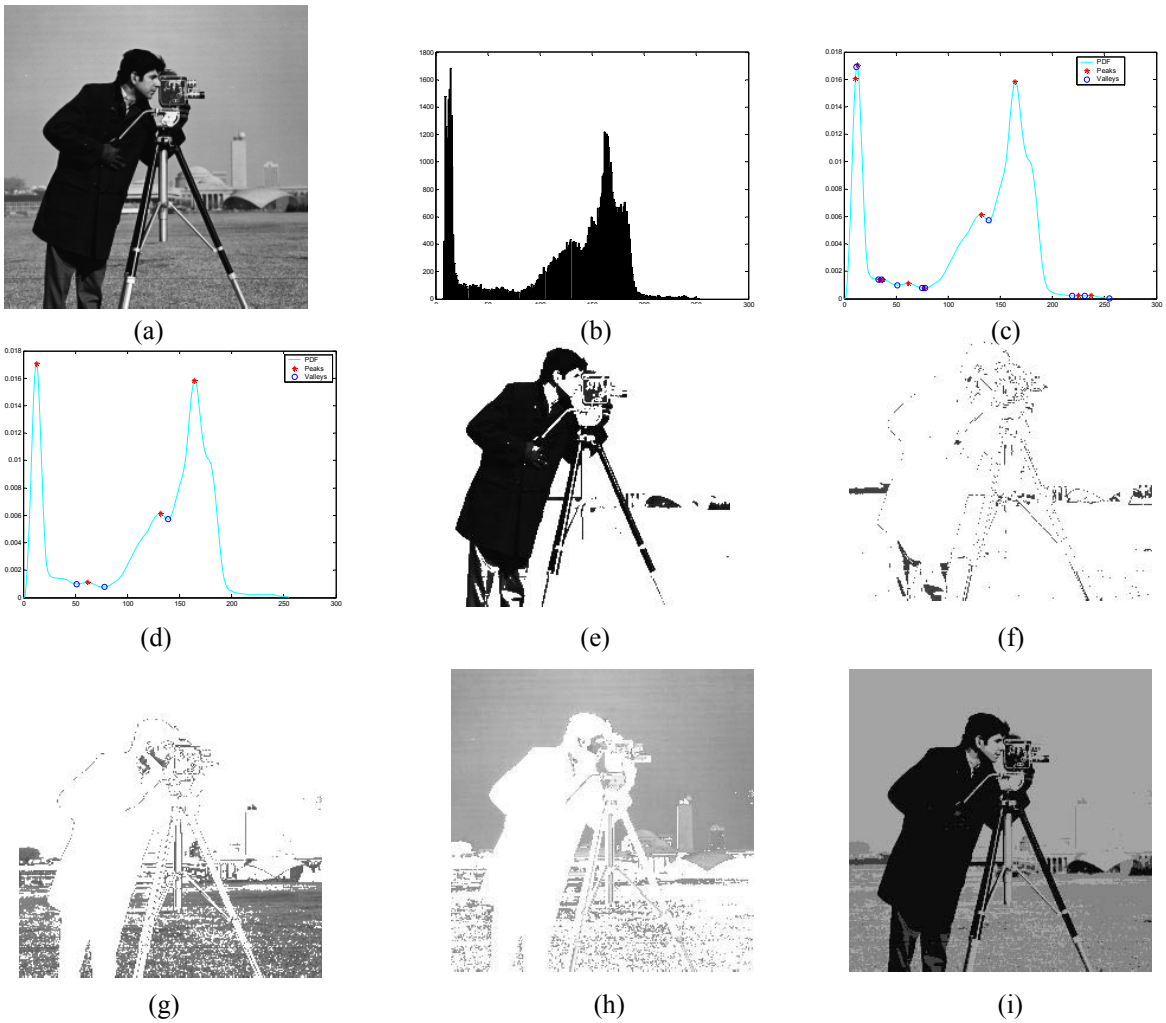
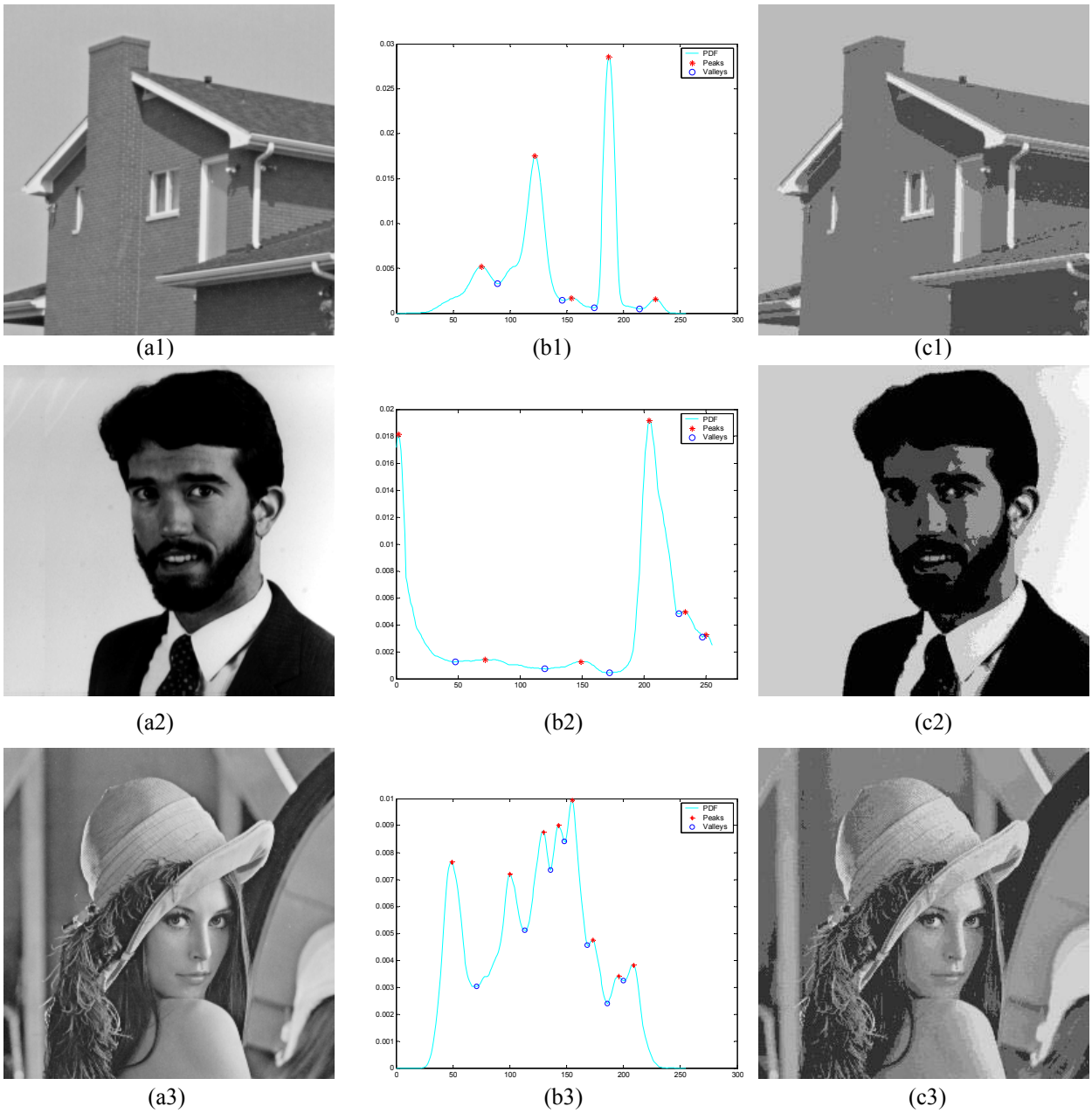


Fig. 3. The segmentation results of the proposed method (h=7). (a) original image of the cameraman; (b) gray-level histogram; (c) peaks and valleys of $\hat{f}(x)$ (calculated by eq.10) before merging; (d) final peaks and valleys; (e)-(h) the resulting segmented images; (i) the final segmented image.

6. Experimental Results.

In this section, we will use several examples to show the performance of the proposed method in segmenting images. Figure 3 demonstrates the segmentation procedures of the proposed method. The original image is shown in figure 3(a) and its gray-level histogram is shown in figure 3(b). Figure 3(c) shows the obtained peaks and valleys before validation. The peaks and valleys after validation are shown in figure 3(d). The segmented images according to the validated valleys are shown in figure (e)-(i).

Before we merged the peaks and valleys, there are ten peaks and ten valleys obtained (figure 3(c)). Near local plateau, there will be some insignificant peaks and valleys. After applying step 5 in the proposed algorithm to validate peaks and valleys, we finally obtained three validated valleys and thus we have four segmented images (figure 3. e-h): (e) corresponds to the cameraman; (g) corresponds to the grassland; and (h) corresponds to the background i.e. the sky. The final segmented image is shown in (i).



(a1) (b1) (c1)

(a2) (b2) (c2)

(a3) (b3) (c3)

Figure 4. The more segmentation results by the proposed method. (a1-a3) the original images; (b1-b3) the final peaks and valleys after validation; (c1-c3) the segmented images.

Figure 4 shows other experimental results. From figure 4, we can see that all the significant peaks and valleys are found, without a priori knowledge about the number of the peaks and valleys.

For lack of space, more experimental results are included in the supplementary material.

The computational speed is efficient. The average time to deal the above images is 0.27 second using MATLAB code on an AMD 800MHz personal computer.

7. Conclusion.

In this paper, the influence of false peak noise on the mean shift method and mean shift valley method is observed. In a 1-D setting, we have given both theoretical analysis and experimental illustration. In this setting, we provide a solution to avoid the influence of false peak noise. The analysis of false peaks in higher dimensional space remains to be addressed.

A novel unsupervised peak-valley sliding algorithm for image segmentation is also presented in this paper. We use the mean shift method to find peaks and the extended mean shift method--mean shift valley to find valleys. The peaks and valleys are alternatively found one by one. After validating the obtained peaks and valleys, we use the validated valleys as density thresholds to segment the image. Several practical problems concerning such a basic scheme were also solved: e.g., oscillations were avoided by adjusting the step size.

It is important to realise that the above analysis shows how to choose h dynamically (dependent upon the data) to avoid false-peak noise. It does not address the issue of choosing h in relation to the smoothing properties in terms of scale-space analysis (determining dominant peaks). Generally speaking, if h is large, the details will be smoothed and the image will be under-segmented; on the other hand, if h is chosen too small, there will be a lot of noise (including peak noise and valley noise) and the image will be over-segmented. One could repeat the analysis with different values of h (still checking for false peaks) to perform a scale-space type analysis. Comaniciu and Meer [5] empirically investigated the influence of the search window radius on the segmentation results. Further suggestions on the choice of the search window radius h can be found in [8]. Rather than implementing a scale space type analysis to remove minor peaks, we used a more heuristic approach. However, these are relatively minor issues compared with the false peak problem.

Our future work is to analyse false peaks in high dimension: including the application of the ideas presented here to segment color images. We believe our work can also be applied in other applications such as MR image analysis, document image analysis, video signal analysis, etc.

References

[1] A. Albiol, L. Torrest, and E.J., Delp, "An Unsupervised Color Image Segmentation Algorithm for Face Detection Applications," Proceedings. 2001 International Conference on Image Processing. Vol: **2**, pp. 681-684, 2001.

[2] H.D. Cheng, Y. Sun, "A Hierarchical Approach to Color Image Segmentation Using Homogeneity," IEEE Trans. Image Processing, **9(12)**, pp. 2071-2082, 2000

[3] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," IEEE Trans. Pattern Analysis and Machine Intelligence, **17(8)**, pp.790-799, 1995.

[4] D. Comaniciu and P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation," in Proceedings of 1997 IEEE Conference on Computer Vision and Pattern Recognition, San Juan, PR, pp.750-755, June 1997.

[5] D. Comaniciu and P. Meer, "Distribution Free Decomposition of Multivariate Data," Pattern Analysis and Applications, **2**, pp.22-30, 1999.

[6] D. Comaniciu and P. Meer, "Mean Shift Analysis and Applications," Proceedings. 7th International Conference on Computer Vision, Kerkyra, Greece, pp.1197-1203, September 1999.

[7] D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift," in Proceedings of 2000 IEEE Conference on Computer Vision and Pattern Recognition, vol 2, pp.142-149, June 2000.

[8] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach towards Feature Space Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, **24**, pp.603-619, 2002.

[9] K. Fukunaga, and L.D. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," IEEE Trans. Info. Theory, vol. IT-21, pp. 32-40, 1975.

[10] Y. Keselman and E. Micheli-Tzanakou, "Extraction and Characterization of Regions of Interest in Biomedical Images," Proceedings. IEEE International Conference on Information Technology Applications in Biomedicine. pp.87-90, 1998.

[11] B.W. Silverman, Density Estimation for Statistics and Data Analysis, London: Chapman and Hall, 1986.

[12] T. Wakahara and K. Ogura, "Extended Mean Shift in Handwriting Clustering," Proceedings, Fourteenth International Conference on Pattern Recognition, **1**, pp. 384-388, 1998.

[13] M. P. Wand and M. Jones, Kernel Smoothing, Chapman & Hall, 1995.

[14] X. Yang and J.Liu, "Unsupervised Texture Segmentation with One-Step Mean Shift and Boundary Markov Random Fields," Pattern Recognition Letters, **22**, pp. 1073-1081, 2001.