16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

EDGE PRESERVING SMOOTHING BY MULTISCALE MODE FILTERING

Vasile Gui

Department of Communications, "Politehnica" University of Timisoara
Bv. V. Parvan No. 2, 300343, Timisoara, Romania
phone: + (40) 256 40 33 06, fax: + (40) 256 40 32 95 , email: vasile.gui@etc.upt.ro
web: www.upt.ro

ABSTRACT

EDGE preserving smoothing is an important step toward image segmentation. Currently, bilateral and mean shift filters are popular solutions to this problem. The multiscale mode filter proposed in this paper can be thought of as a generalization of both, working successively at a number of scales. As a result of the multiscale approach, the filter avoids being trapped into spurious local maxima, in favour of more significant ones. Compared with the conventional mean shift filter, the multiscale mean shift filter produces smoother images and restores blurred edges more effectively. The filter can cope well with images corrupted by heavy noise.

1. INTRODUCTION

Edge preserving image smoothing plays an important role in image processing and computer vision. Often image smoothing is a pre-processing step toward image segmentation. Low-pass linear filters assume local image constancy or slow variation and uncorrelated noise. However these assumptions are invalid in the presence of edges and/or non-significant local image detail supposed to be rejected by the filter. Since the likelihood of having similar image value with the central pixel in an image window decreases with the distance to the centre, linear smoothing filters weight less the pixels as their distance to the centre increases. While this is a step in a right direction, resulting in less edge blurring, noise rejection and smoothing effects are also diminished. A better compromise can be obtained in a nonlinear image smoothing framework.

Among the oldest ideas in edge preserving image smoothing methods is due to Graham [1]. Pixels corrupted by impulse noise were detected and replaced by an estimate based on local average. Since then, several solutions have been proposed to limit the effect of "untypical" or "outlier" samples in the filtering window. Kuwahara [2] and Nagao-Matsuiama [3] avoid averaging across edges by searching for locally smooth neighbourhoods of image pixels. More recently, in anisotropic diffusion [4][5], iterative averaging in neighbourhoods, depending on measured local variation, is used in order to avoid edge blurring. The bilateral filter proposed by Tomasi and Manduchi [6] treats space domain (pixel *coordinates*) and range domain (pixel *values*) in a unified framework. Weighted averaging is performed in the joint range-space domain, with weights depending on distance in the joint domain between pixels in the image and the currently processed pixel. The weighting function proposed is the product of two functions, depending on space and respectively range domain distance, but the functions play a symmetrical role. A very similar solution is the local mode filter proposed by Weijer and Boomgaard [7], where the connection with multidimensional histogram's local modes is made. A more general solution, with theoretically rigorous demonstration of the convergence to a local density mode is the mean shift filter [8]. The mean shift algorithm convergence to local maxima of a probability density function makes it suitable for many computer vision tasks, such as segmentation [8], tracking [9] or background estimation [10]. Mean shift is an iterative solution for local mode location estimation, based on kernel probability density estimation and a gradient ascent optimization scheme. Theoretical links between the bilateral filter, the mean shift filter and nonlinear diffusion can be found in Barrash [11].

A critical step in the design of any mode filter is the choice of the scale or filter bandwidth. Scale estimation is a classical problem in nonparametric density estimation. Data driven solutions dominate in the robust statistics literature, while problem driven scale selection or mixed solutions can be found in image processing papers. Adaptive scale selection is described in [12]. The variable bandwidth mean shift filter used in [12] addresses the problem by adapting the bandwidth parameters to the estimated level of noise. While the approach is undoubtedly better than a constant bandwidth solution, excessive peak smoothing and insufficient tail rejection still may be objected. To overcome this remaining draw back, we propose the multiple scale mode filter (MSMF). It can be designed to work either as a multiple scale bilateral filter or as a multiple scale mean shift filter. Multiscale methods have been used in data processing for many reasons. Among them, processing speed and solution stability are the most important.

2. RELATED WORK

Bilateral and mean shift filters work in a joint domain, consisting of spatial data and range data. For static images, the spatial domain is composed by pixel coordinates vectors:

$$\mathbf{x}_s = \begin{bmatrix} x \\ y \end{bmatrix}.$$

The range information may be intensity level, colour vector or another vector, measuring some features at a specific spatial coordinate. For colour images using RGB space, the range vector may be expressed as:

$$\mathbf{x}_r = f(\mathbf{x}_s) = \begin{bmatrix} r(\mathbf{x}_s) \\ g(\mathbf{x}_s) \\ b(\mathbf{x}_s) \end{bmatrix}.$$

In this case, a joint domain data has the form:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_s \\ f(\mathbf{x}_s) \end{bmatrix} = \begin{bmatrix} x \\ y \\ r(x,y) \\ g(x,y) \\ b(x,y) \end{bmatrix}.$$

Since generally space and range data have different scales, data is supposed to be conveniently normalized prior to joint domain representation. A 2D image consisting of N pixels is represented by the set of 5D vectors: $\{\mathbf{x}_i\}$, $i = 1,2,\ldots,N$. Let $H(\mathbf{x}_1,\mathbf{x}_2)$ be a function measuring the similarity of two vectors, $\mathbf{x}_1$, $\mathbf{x}_2$. The bilateral filter response at pixel $\mathbf{x}_c$, is defined by an equation of the form:

$$\mathbf{y}_{cr} = F(\mathbf{x}_{cs}) = \frac{\sum_{i=1}^{N} \mathbf{x}_{ir} H(\mathbf{x}_c, \mathbf{x}_i)}{\sum_{i=1}^{N} H(\mathbf{x}_c, \mathbf{x}_i)} . \qquad (1)$$

Clearly, $\mathbf{y}_{cr}$, the output range vector at location $\mathbf{x}_{cs}$ is a weighted sum of input image range vectors, $\mathbf{x}_{ir}$, with weights defined by the similarity function $H()$. The sum at the denominator is a normalization factor needed to make weights add up to 1 in order to preserve the mean of each component of the range data. To have a high influence on the currently computed output image, an input image pixel needs to be similar with the central pixel, in both location and value. This is why bilateral filters are able to effectively smooth the image without blurring edges or cutting out detail with high contrast but low spatial extent.

The mean shift filter relies on the mean shift algorithm proposed by Fukunaga and Hostetler [13] to find local modes of the estimated probability density in a data set. Starting from a data sample, $\mathbf{x}$, the algorithm successively estimates the mean shift vector, pointing in the direction of the estimated density gradient and makes a step in the direction of the gradient. Since the gradient vanishes at local maxima, the algorithm converges when a local density maximum is reached. Assuming a radially symmetric kernel, the following steps are iterated for each pixel, $\mathbf{x}_c$:

1. Set the current result as the current input pixel: $\mathbf{y}_0 = \mathbf{x}_c$.
2. Compute the next value of the current result:

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

$j = 1,2,\ldots$

until convergence has been reached, that is the mean shift vector norm drops under a small threshold:

$$\left\|\mathbf{y}_{j+1} - \mathbf{y}_j\right\| = \|\mathbf{m}\| < \varepsilon .$$

The result of filtering $\mathbf{x}_c$ is $\mathbf{y}_{j+1}$. Only the range information of the result is stored in the output image. In equation (2), $g(x) = -k'(x)$, and $k(x)$ is the kernel profile used to estimate the probability density:

$$\hat{p}_{k,h}(\mathbf{x}) = \frac{c_{k,h}}{N} \sum_{i=1}^{N} k\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right). \qquad (3)$$

The profile, $k()$, acts as an interpolation function needed to generate a continuous density estimate from a finite set of data samples. It can be shown that the density estimate obtained by using equation (3) is the convolution of the real density with the kernel function [14]. The scale factor modulates the shape of the kernel function and controls the ammount of smoothing.

A mean shift filter replaces each pixel's range data with the range data of the mean shift algorithm convergence point, obtained with iterations starting from that pixel. The result corresponds to a local maximum of the probability density of the data, estimated with a certain kernel. The profile $k(x)$ is called the shadow profile of $g(x)$. If $k(x)$ is Gaussian, then $g(x)$ is also Gaussian with the same scale. More generally, the scale is defined by a $d \times d$ bandwidth matrix. The radially symmetric case considered in this paper corresponds to a bandwidth matrix of the form $h\mathbf{I}$, where $\mathbf{I}$ is the unit matrix. Mean shift filter is closely related to mean shift clustering and segmentation. Indeed, by associating all pixels converging to the same mode into the same class or segment, the mean shift agorithm can be used for image segmentation. However, a postprocessing stage is needed to detect and link close modes.

The relationship between the mean shift filter and the bilateral filter can be easily understood by comparing the equations (1) and (2). If the similarity function of the bilateral filter is set to be equal to the kernel profile of the mean shift filter, that is $H(\mathbf{x}_i, \mathbf{x}_j) = g(\|\mathbf{x}_i - \mathbf{x}_j\|^2/h^2)$, the bilateral filter returns the result of the first iteration of the corresponding mean shift filter. Therefore, the mean shift filter is more general and its result corresponds to the location of a maximum point of the estimated probability density function of the image data.

3. MULTI SCALE MODE FILTER (MSMF)

Let $\{h_j\}, j = 1,2,\ldots, J$ be a number of scales, associated with a spherical kernel profile $g()$. The multiscale mode filter of an image with N pixels is defined by the algorithm shown in Figure 1.

For each image pixel, $\mathbf{x}_c$, do:

1. Set the current result as the current input pixel:

$$\mathbf{y}_0 = \mathbf{x}_c.$$

2. for j = 0 to J-1 compute

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^{N} \mathbf{x}_i g\left(\left\| \frac{\mathbf{y}_j - \mathbf{x}_i}{h_j} \right\|^2\right)}{\sum_{i=1}^{N} g\left(\left\| \frac{\mathbf{y}_j - \mathbf{x}_i}{h_j} \right\|^2\right)}, \qquad (4)$$

3. Set the result as $\mathbf{y}_J$.

Figure 1 – Multiscale mode filter algorithm

According to the value of J and the choice of scales h_j, several multiscale mode filters can be designed. The case $J = 1$ corresponds to a conventional, single scale, bilateral filter. The case $h_j = h_0$, for all j, with J sufficiently high, corresponds to the conventional mean shift filter. A multiscale bilateral filter can be obtained by freezing the space coordinates in equation (4). A monotonically decreasing set of scales $h_{j+1} < h_j$ for any $j < J$ is the main case motivating the proposed generalization. Our purpose in doing so is twofold:

- to reduce the influence of the starting point in the initial steps of the algorithm in favor of more confidence given to the context
- to avoid the algorithm being trapped into spurious local maxima of the density gradient

Both mean shift and bilateral filters rely heavily on the starting point in the joint range-space domain. Because of this, when the noise level is increased, the output may be severely offset. As experiments with real data have shown, given a desired scale of analysis, the probability density function often has several local maxima close to one another. Since the mean shift is a gradient ascent type of algorithm, it may be trapped in such a spurious density maximum point. This event is more likely to happen at low signal to noise ratios. If a large scale is used to smooth the estimated pdf, spurious local modes can be removed at the expense of shifting the locations of the maxima, when the distributions are not symmetrical. This effect is illustrated in Fig. 2 for a 1D blurred impulse signal with added noise. However the filter restores the edges of the signal (avoid spurious local modes) and in the same time accurately reconstructs the signal levels (that is accurately finds the locations of the two maxima). The large scale mode is found first. Then the mode location is adjusted in the second step of the MSMF, working at a finer scale. A closer, spurious mode would have not been reached by searching directly at the final scale.

The highest scale, h_1, defines the degree of smoothing of the density field needed to clean out spurious local maxima, while the final scale of analysis, h_J, is supposed to be obtained by one of the many existing techniques [10].

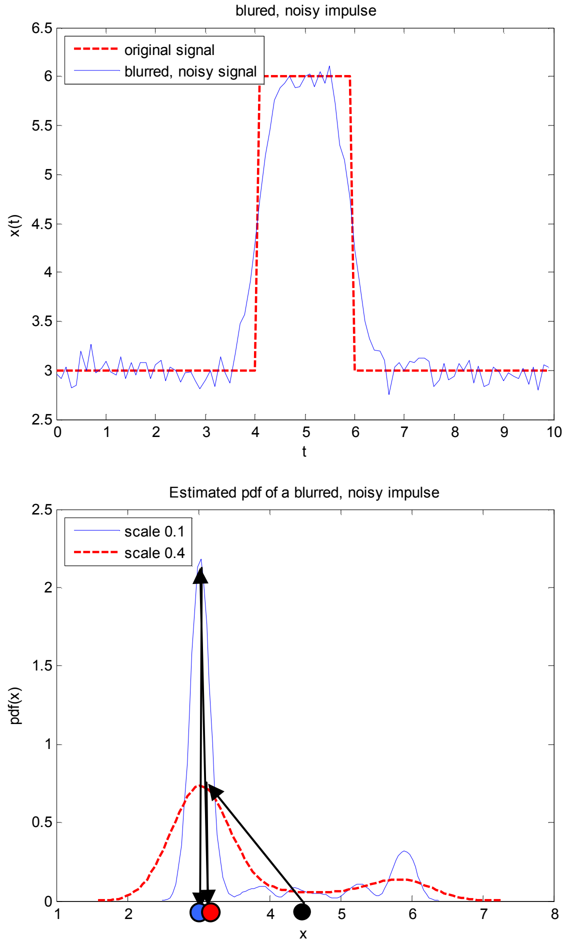
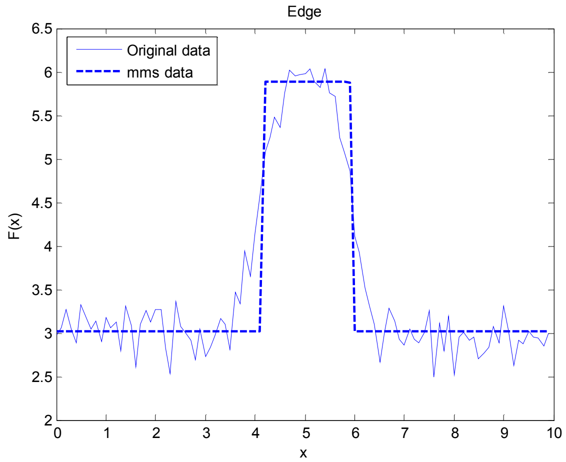


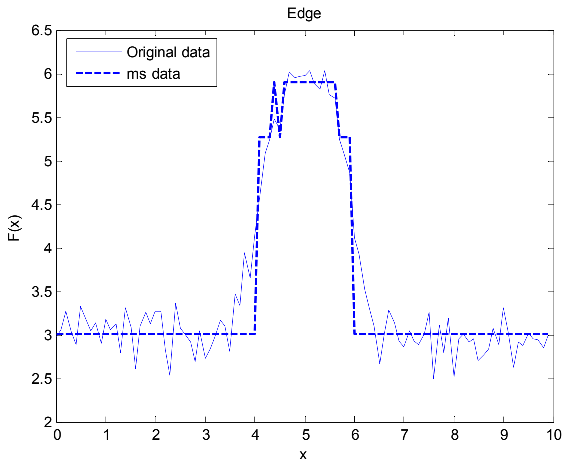
Figure 2 – Multiscale mode finding on 1D edge data. Up: noisy impulse. Down: Large scale pdf mode (red dot) is found first, then lower scale mode (blue dot) is precisely reached. Spurious modes around starting point (black dot) avoided.

4. EXPERIMENTAL RESULTS AND DISCUSSION

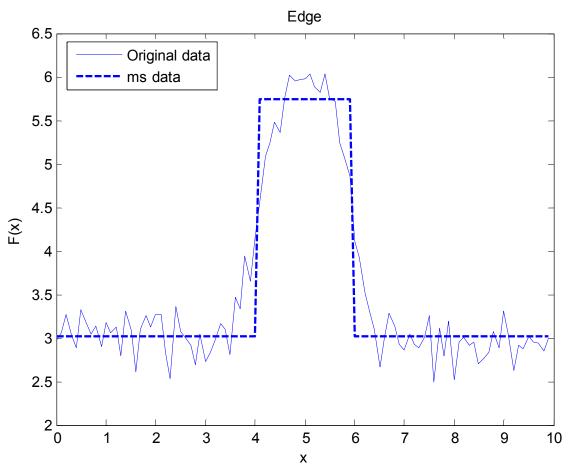
We first illustrate the results of the MSMF on the 1D blurred and noisy, rectangular signal. The original signal was a sequence of 100 samples, with 80 samples on level 3 and 20 samples on level 6. The signal was blurred by convolution with a binomial filter of length 33. A sequence of Gauss distributed white noise with zero mean and a standard deviation of 0.2 was added to the signal, then a conventional mean shift filter and a MSMF were applied. The MSMF was run with $h_1 = 1$, $h_J = 0.2$ and $J = 10$, while the conventional mean shift filter was run with scales h, $2h$ and $3h$. The results are shown in Fig. 3. Clearly, the MSMF is able to restore the original signal, as seen in Fig. 3a, while the conventional mean shift filter with the same final scale is not, as seen in Fig. 3b. It removes the noise equally well, but is unable to restore the true edge, because of the existence of multiple maxima on the blurred edge. When the scale of the mean shift filter is increased progressively to 2*h* and 3*h*, multiple maxima are also eliminated, the true edge is well detected, but the detected amplitude is progressively diminished. This happens because of the larger tails of the Gaussian kernel, enabling higher influence of samples not belonging to the detected mode.



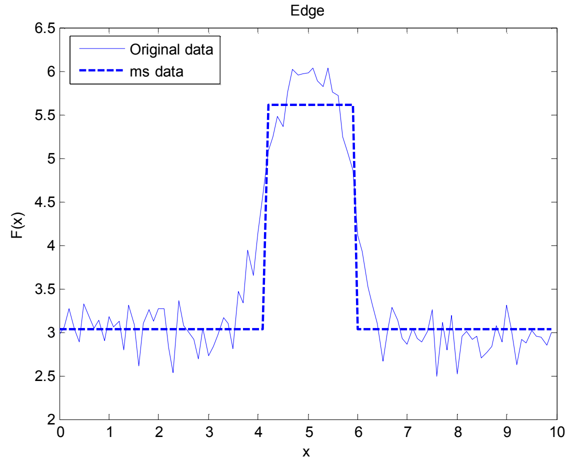
a)



b)



c)



d)

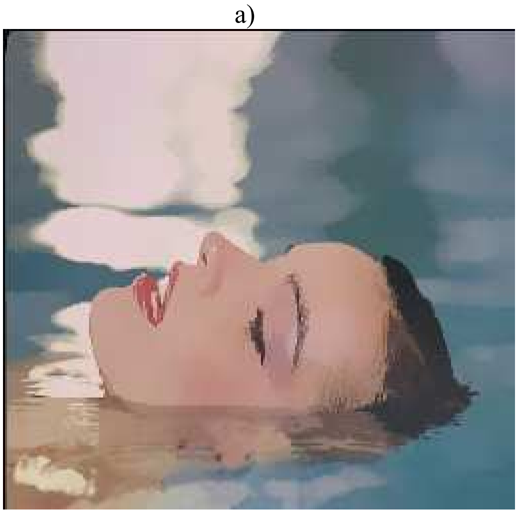
Figure 3 –1D blurred noisy edge response: a) –MSMF filter response, with final scale $h = 0.2$; b), c) and d) – multiscale mode filter response with scales *h*, 2*h* and 3*h* respectively.

We applied the MSMF on images, as a pre-processing step toward segmentation. The original image, shown in Fig. 4a) contains a moderate amount of noise and some blurred edges. The response of the mean shift filter with scale $h = 17$ on the original image is shown in figure 4b), while the response of the MSMF with $h_1 = 33$, $h_J = 17$, $J = 8$ and equal scale steps is shown in figure 4c). A comparison of the images in figure 4b) and 4c) reveals that the MSMF produces a smoother output, with more regular-shaped colour patches and restores edges better than the mean shift filter working at the same final scale.

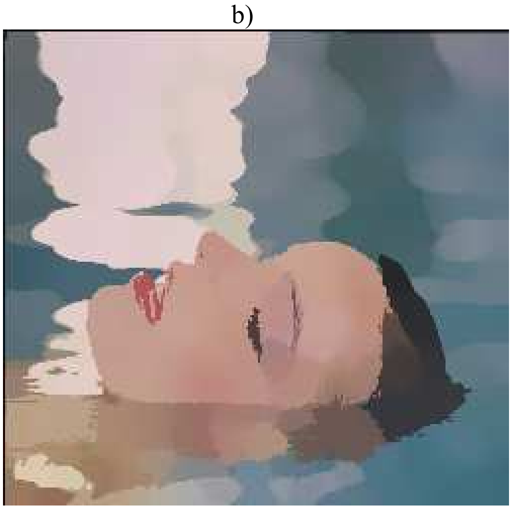
In spite of being more general than the mean shift filter, the computational complexity of the MSMF is not necessarily higher. Unlike in the case of the mean shift filter, where the number of steps needed to obtain convergence is unknown and varies from pixel to pixel, the computational complexity of the MSMF exclusively depends on the number of steps and the kernel size, related directly to the spatial scale parameters. While for the MSMF convergence to a local mode is not implicit, a sufficient number of steps, of the order of magnitude of 10, lead to asymptotically stable results when the scale is frozen. If the convergence is an issue, the filter can be run at the final scale until convergence is obtained.

The experiments included in this work prove the usefulness of the proposed MSMF in some cases. Since the filter is a generalization of the mean shift and bilateral filters it is no surprise that, in some cases, it can outperform the particular filters. However, the proper design of the MSMF is a matter of further study. What is the best filtering scenario? Can the same solution be obtained by the mean shift filter with another scale parameter in a certain application? Instead of one scale parameter there are many more in the proposed approach. Moreover, if the smoothing kernels use non-diagonal bandwidth matrices, the number of scale parameters and choices for the MSMF if further increased. Since the invention of nonparametric kernel density estimation method, after several decades, the scale selection problem is still a subject of active research. In our view, this happens because the relevance of the currently used scale selection methods to the application at hand cannot be easily assessed.

a)



b)



c)

Figure 4 – Comparative results of the mean shift filter and the multiscale mode filter: a) original image; b) mean shift filtered image; c) result of the MSMF

5. CONCLUSIONS

In this paper we proposed a multiscale mode filter, as a generalization of the bilateral and mean shift filters. The filter can be used for edge preserving smoothing or as a preprocessing step in image segmentation. The results were compared with the results of the mean shift filter. At the same analysis scale, the proposed MSMF generates smoother images and sharper edges.

REFERENCES

[1] R.E. Graham, "Snow-removal - a noise-stripping process for picture signals", *IRE Transactions on Information Theory*, 8, pp. 129-144, 1961.

[2] M. Kuwahara, K. Hachimura, S. Eiho, and M. Kinoshita, “Processing of RI-angiocardiographic images”, in *Digital Processing of Biomedical Images*, K. Preston, Jr. and M. Onoe, Editors. New York: Plenum, 1976. pp. 187-202.

[3] M. Nagao and T. Matsuyama, "Edge preserving smoothing", *Computer Graphics and Image Processing*, vol. 9, no. 4, pp. 394-407, Apr. 1979.

[4] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion”, *IEEE Trans.*, *PAMI-12(7), pp.* 629–639, 1990.

[5] G. Sapiro and D. L. Ringach. “Anisotropic diffusion of color images”, *Proc. SPIE*, 2657:471–382, 1996.

[6] C. Tomasi, R. Manduchi, “Bilateral filtering for gray and color images", *Int'l Conf. Comp. Vis., Bombay, India*, pp. 839-846, 1998.

[7] J.Van de Weijer and R. van den Boomgaard, “Local mode filtering”, in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Vol. 2*, pp.428-433, 2001.

[8] D.Comaniciu, P.Meer, “Mean shift: a robust approach toward feature space analysis”, *IEEE Trans. Pattern Recognition & Machine Intelligence,Vol.24,No.5*,pp.603-619, 2002.

[9] D.Comaniciu, V. Ramesh, P.Meer, “Kernel-based object tracking”, *IEEE Trans. Pattern Recognition & Machine Intelligence,Vol.25,No.5*,pp.564-577, 2003.

[10] B. Han, D. Comaniciu, and L.Davis, "Sequential kernel density approximation through mode propagation: applications to background modeling,“ Proc. ACCV -Asian Conf. on Computer Vision, 2004.

[11] D.Barash, “A Fundamental relationship between bilateral filtering, adaptive smoothing and the nonlinear diffusion equation”, *IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.24, No.6*, pp.844-847, 2002.

[12] D. Comaniciu, V. Ramesh, P. Meer, “The variable bandwidth mean shift and data-driven scale selection”, *IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.25, No.2*, pp.281-288, 2003.

[13] K. Fukunaga and L. Hostetler, “The estimation of the gradient of a density function, with applications in pattern recognition”, *IEEE Trans. Info. Theory*, IT-21, pp32-40, 1975.

[14] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*. 2nd ed., John Wiley & Sons, New York, NY, 2001.