ML NONLINEAR SMOOTHING FOR IMAGE SEGMENTATION AND ITS RELATIONSHIP TO THE MEAN SHIFT

Andy Backhouse [a], *Irene Y. H. Gu* [a], *Tiesheng Wang* [b]

[a]Dept. of Signals and Systems, Chalmers University of Technology, Sweden
[b]Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong Univ., China

ABSTRACT

This paper addresses the issues of nonlinear edge-preserving image smoothing and segmentation. A ML-based approach is proposed which uses an iterative algorithm to solve the problem. First, assumptions about segments are made by describing the joint probability distribution of pixel positions and colours within segments. Based on these assumptions, an optimal smoothing algorithm is derived under the ML condition. By studying the derived algorithm, we show that the solution is related to a two-stage mean shift which is separated in space and range. This novel ML-based approach takes a new kernel function. Experiments have been conducted on a range of images to smooth and segment them. Visual results and evaluations with 2 objective criteria have shown that the proposed method has led to improved results which suffer from less over-segmentation than the standard mean-shift.

Index Terms— Maximum Likelihood, Mean-Shift Procedure, Image Smoothing, Image Segmentation

1. INTRODUCTION

Image segmentation has been widely studied [3]. It generates a set of partitions in an image which consist of homogeneous regions. Different segmentation approaches fall into categories such as pixel, region, content, edge, object and semantic-based methods. Stochastic modelling techniques are commonly used in the segmentation of textures and complex images [1, 2]. Although they are very robust, these techniques often suffer from being computationally intensive. Other techniques such as the mean shift, bilateral filter and nonlinear diffusion are frequently employed due to their ability to smooth an image while preserving edges [5, 4, 3]. These techniques, which can all be related through a common framework [9], are robust yet less computational intensive than most stochastic methods. However, they are sensitive to kernel bandwidth selection.

In this paper, we propose a novel nonlinear smoothing and segmentation algorithm, which is algorithmically similar to the mean-shift. The most significant difference between the two lies in their derivation. The proposed method is derived from an analysis of the joint-probability that two pixels are contained in the same segment. The derivation leads to a maximum likelihood solution for the mean colour in each segment. This derivation makes the connection between the statistical assumptions and parameters selection much clearer than for the standard mean-shift.

Algorithmically, the proposed algorithm consists of two consecutive weighted mean-shift algorithms. Compared to the standard mean-shift, both algorithms preserve edges. However the proposed algorithm smooths the inside of segments while the mean-shift algorithm often over-segments them. This smoothing is beneficial because it makes the merging of segments easier to perform.

The paper is organized as follows. In Section 2, a maximum likelihood (ML) approach is formulated which derives an estimate of the mean-segment colour for each pixel based on the probability that any two pixels belong to the same segment. In Section 3, the probability that two pixels belong to the same segment is derived. Next, in Section 4, after a short review of the mean shift filter, the similarity and difference between the ML-based approach and the mean shift are discussed. Section 5 includes some experimental results and comparisons. Finally, conclusions are given in Section 6.

2. MAXIMUM LIKELIHOOD ANALYSIS FOR IMAGE SEGMENTATION

In this section, the image segmentation problem is formulated in terms of a maximum likelihood problem. For every pixel, we derive the probability that its encapsulating segment has a given mean colour. We derive this probability subject to the spatial and colour characteristics of all other pixels. Based on this, it is possible to find the colour that maximises the likelihood of the mean segment colour.

In the following text, the spatial position of pixel i shall be denoted by $\mathbf{s}_i$ and its colour by $\mathbf{c}_i$. The segment containing the ith pixel shall be denoted $\mathcal{X}_i$. Further, any two given segments, $\mathcal{X}_i$ and $\mathcal{X}_j$, are either identical or disjoint. We assume that a pixel j is contained in segment $\mathcal{X}_i$ with probability $P(j \in \mathcal{X}_i|\mathbf{s}_j)$. This shall be denoted $P(\mathbf{s}_j)$ and is given by

$$P(\mathbf{s}_j) \propto \exp\left\{-\left\|\mathbf{s}_j - \bar{\mathbf{s}}_i\right\|^2 / \sigma_s^2\right\},$$

where $\bar{\mathbf{s}}_i$ is the centre of segment $\mathcal{X}_i$ and is assumed known. We further assume that the colours of all pixels within a segment are i.i.d. Gaussian distributed around the mean colour of the segment, and the colours of pixels in different segments are white and uniformly distributed in the colour space. This is equivalent to having no prior information about the colour values of pixels. That is, there are two cases:

$$\begin{array}{lll} j \in \mathcal{X}_i \text{ and } & \mathbf{c}_j = \bar{\mathbf{c}}_i + \mathbf{w}_j, & \mathbf{w}_j \sim \mathcal{N}(0, \sigma_c^2 \mathbf{I}) \\ j \notin \mathcal{X}_i \text{ and } & \mathbf{c}_j = \mathbf{d}_j, & \mathbf{d}_j \sim \frac{1}{255^3} \end{array}$$

Based on the above two cases, the joint probability of colour and position of each pixel can be described as follows:

$$P(\mathbf{c}_j|\mathbf{s}_j, \bar{\mathbf{c}}_i) \propto \frac{P(\mathbf{s}_j)}{(2\pi\sigma_c^2)^{3/2}} e^{-\|\mathbf{c}_j - \bar{\mathbf{c}}_i\|^2/\sigma_c^2} + \frac{1}{255^3}(1 - P(\mathbf{s}_j))$$

Hence, the joint probability for all pixels $j \in \mathcal{X}_i$ (with mean segment colour $\bar{\mathbf{c}}_i$) is given by

$$P(\mathbf{c}_1 \cdots \mathbf{c}_N|\mathbf{s}_1, \cdots \mathbf{s}_N, \bar{\mathbf{c}}_i) \propto \prod_j P(\mathbf{c}_j|\mathbf{s}_j, \bar{\mathbf{c}}_i) \qquad (2.1)$$

The best estimate $\hat{\mathbf{c}}_i$ of $\bar{\mathbf{c}}_i$, is the maximum likelihood solution to the log-likelihood of (2.1), which is given by the equation

$$\hat{\mathbf{c}}_i = \arg\max_{\bar{\mathbf{c}}_i} \ln P(c_1 \cdots \mathbf{c}_N|\mathbf{s}_1, \cdots \mathbf{s}_N, \bar{\mathbf{c}}_i).$$

Applying the logarithm to (2.1), taking the derivative with respect to $\bar{\mathbf{c}}_i$, and setting it to zero, yields the following

$$\begin{aligned} &\frac{\partial}{\partial \bar{\mathbf{c}}_i} \left(\ln P(c_1 \ldots c_N|s_1 \ldots s_N, \bar{\mathbf{c}}_i)\right) \\ &= \sum_j \frac{255^3(\mathbf{c}_j - \bar{\mathbf{c}}_i)P(\mathbf{s}_j)e^{-\|\mathbf{c}_j - \bar{\mathbf{c}}_i\|/\sigma_c^2}}{255^3 P(\mathbf{s}_j)e^{-\|\mathbf{c}_j - \bar{\mathbf{c}}_i\|^2/\sigma_c^2} + (2\pi\sigma^2)^{3/2}(1 - P(s_i))} = 0 \end{aligned}$$

which is equivalent to

$$\hat{\mathbf{c}}_i = \frac{\sum_j c_j K(\mathbf{c}_j - \hat{\mathbf{c}}_i, P(\mathbf{s}_i))}{\sum_j K(\mathbf{c}_j - \hat{\mathbf{c}}_i, P(\mathbf{s}_i))}, \qquad (2.2)$$

where $K(\mathbf{u}, v) = \sum_j \frac{255^3 \exp\left(-\|\mathbf{u}\|^2\sigma^{-2}\right)}{255^3 \exp\left(\|\mathbf{u}\|^2\sigma^{-2}\right) + (2\pi)^{3/2}\sigma^3(\frac{1}{v} - 1)}$. It is worth noting that (2.2) has a remarkable resemblance to the mean-shift expression. This shall be developed further in section 4.

3. MAXIMUM LOG-LIKELIHOOD ESTIMATE FOR THE SPATIAL CENTRE

In Section 2, it is assumed that the centre $\bar{\mathbf{s}}_i$ of segment $\mathcal{X}_i$ is known in advance so that (2.2) can be computed. To estimate $\bar{\mathbf{s}}_i$, colour information will be used to find the most likely candidate. This will rectify the assumption made in section 2, that colour differences between pixels do not affect $P(\mathbf{s}_j)$. Mathematically, the derivation in this section mirrors the previous section. First, we assume that the probability that a pixel j is in the segment $\mathcal{X}_i$ subject to its colour, is given by

$$P(j \in \mathcal{X}_i|\mathbf{c}_j) = \exp\left\{-\|\mathbf{c}_j - \bar{\mathbf{c}}_i\|^2/\sigma_c^2\right\},$$

where $\bar{\mathbf{c}}_i$ is the mean colour of segment $\mathcal{X}_i$ and is assumed known. We denote $P(j \in \mathcal{X}_i|\mathbf{c}_j)$ by $P(\mathbf{c}_j)$ for simplification. The probability that $\bar{\mathbf{s}}$ is the centre of $\mathcal{X}_i$ is given by

$$P(\bar{\mathbf{s}}) = \sum_{\mathcal{S}} \delta\left(\bar{\mathbf{s}} = \frac{\sum_{j \in \mathcal{S}} \mathbf{s}_j}{\sum_{j \in \mathcal{S}} 1}\right) P(\mathcal{S} = \mathcal{X}_i),$$

where $\sum_{\mathcal{S}}$ is the sum over all possible connected sets. To obtain the solution to the above equation, an approximation is made by limiting the possible $\mathcal{S}$ to spherical shapes. In such a case, the above equation can be reasonably simplified since any sphere centered at $\mathbf{s}_i$ is guaranteed to satisfy the delta function. Let a sphere centered at $\mathbf{s}_i$ with radius r be denoted $\mathcal{B}(\mathbf{s}_i, r)$, then it follows

$$P(\bar{\mathbf{s}}) \approx \int_r P(\mathcal{B}(\mathbf{s}_i, r)). \qquad (3.1)$$

Assuming all pixels can be treated separately and that probability decreases as the size of the segment increases, then

$$P(\mathcal{B}(\bar{\mathbf{s}}, r)) = P(r) \prod_{j \in \mathcal{B}(\hat{\mathbf{s}}, r)} P(\mathbf{c}_j) \prod_{j \notin \mathcal{B}(\hat{\mathbf{s}}, r)} (1 - P(\mathbf{c}_j))$$

Set $P(r) = \exp\{-r/\sigma_s^2\}$, (3.1) becomes

$$P(\bar{\mathbf{s}}) = \int_0^\infty e^{-r/\sigma_s^2} \prod_{j \in B(\bar{\mathbf{s}}, r)} P(\mathbf{c}_j) \prod_{j \notin B(\bar{\mathbf{s}}, r)} (1 - P(\mathbf{c}_j)) dr.$$

Taking the logarithm, we get

$$\begin{aligned} \ln P(\bar{\mathbf{s}}) &= \sum_j \int_{r < \|\mathbf{s}_j - \hat{\mathbf{s}}_i\|^2} P(\mathbf{c}_j) e^{-r/\sigma_s^2} dr \\ &\quad + \int_{r > \|\mathbf{s}_j - \bar{\mathbf{s}}_i\|^2} (1 - P(\mathbf{c}_j)) e^{-r/\sigma_s^2} dr \\ &= \alpha + \beta \sum_i P(\mathbf{c}_j) e^{-\|\bar{\mathbf{s}}_i - \mathbf{s}_j\|^2/\sigma_s^2} \end{aligned}$$

where α and β are constants. Taking the derivative and setting this equal to zero, yields that the ML estimate $\hat{\mathbf{s}}_i$ satisfies $\sum_j (\hat{\mathbf{s}}_i - \mathbf{s}_j)\beta P(\mathbf{c}_j) e^{-\|\hat{\mathbf{s}}_i - \mathbf{s}_j\|^2/\sigma_s^2} = 0$, which implies

$$\hat{\mathbf{s}}_i = \frac{\sum_j \mathbf{s}_j K_s(\hat{\mathbf{s}}_i - \mathbf{s}_j, P(\mathbf{c}_j))}{\sum_j K_s(\hat{\mathbf{s}}_i - \mathbf{s}_j, P(\mathbf{c}_j))} \qquad (3.2)$$

where $K_s(\mathbf{u}, v) = v e^{-\|\mathbf{u}\|^2/\sigma_s^2}$.

4. ML-BASED SEGMENTATION: SIMILARITY AND DIFFERENCE TO A STANDARD MEAN SHIFT

In this section, we show that the proposed ML-based image smoothing is similar in spirit to the mean shift. Mean-shift seeks local maxima (modes) from the kernel-based density estimate. Given a set of data $\{\mathbf{x}_i, i = 1, ..., N\}$, the kernel density estimate of $\mathbf{x}$ is given by

$$p_K(\mathbf{x}) = \frac{1}{N\mathbf{H}} \sum_i K(\mathbf{H}^{-1} d(\mathbf{x}, \mathbf{x}_i, \mathbf{H})) \tag{4.1}$$

The kernel $K(\cdot)$ is defined here as a radially-symmetric positive integrable function. Two commonly used kernels are the Gaussian kernel and the Epanechnikov kernel. The shape of the kernel combined with the bandwidth matrix $\mathbf{H}$ determines how much, data close to $\mathbf{x}$, affect the density estimate. The mean-shift [4] is designed to find the local maxima in the kernel-based estimate of the pdf. Given a kernel K, there exist a shadow kernel G such that the iterated process

$$x^{n+1} = \frac{\sum_{i=1}^{N} \mathbf{x}_i G(\|\frac{\mathbf{x}^n - \mathbf{x}_i}{h}\|^2)}{\sum_{i=1}^{N} G(\|\frac{\mathbf{x}^n - \mathbf{x}_i}{h}\|)^2} \tag{4.2}$$

is guaranteed to converge to a local maximum of the density estimate. The mean-shift is defined as $m(\mathbf{x}^n) = \mathbf{x}^{n+1} - \mathbf{x}^n$ and tends to 0, when $\mathbf{x}^n$ reaches the local maximum. In image segmentation, the data set is chosen as a 5D feature set which is defined as the vector $\mathbf{x}_i = [\mathbf{s}_i, \mathbf{c}_i]^T$ that includes both the position and the value of image pixels. The mean-shift is run on all pixels. Once a local maximum is found, the original pixel colour is replaced by the converged colour value.

The proposed algorithm is similar to the mean-shift algorithm. First, an estimate of the centre $\bar{\mathbf{s}}_i$ of each segment $\mathcal{X}_i$ is derived assuming that the the mean segment colour $\bar{\mathbf{c}}_i$ is known and the probability that j is contained in $\mathcal{X}_i$ is only dependent on its proximity in colour (see Section 4). The centre is found with the use of a mean-shift algorithm (Equation (3.2)). The kernel is a weighted Gaussian kernel. Next, a better estimate of the segment colour $\bar{\mathbf{c}}_i$ is obtained by maximising the likelihood function for the colour (see Section 3). This is achieved with a mean-shift algorithm (Equation (2.2)), which is a new type of kernel. The algorithm is run for all pixels and then repeated until the colours converge. Usually 5-6 iterations are sufficient. The algorithm is summarized in Table 1.

Although the two methods are similar, the ML-based approach has less over-segmentation. This is because the two-stage mean-shift algorithm used by the proposed method are both weighted. Equation (3.2) is weighted by the initial estimate of the mean-segment colour and therefore points which are close, but differ in colour, will not converge to exactly the same spatial position. The same holds for the second part of the algorithm. This is in contrast to the standard mean-shift where close points are guaranteed to converge to the same point and therefore over-segmentation can occur.

5. EXPERIMENTAL RESULTS

Experiments have been conducted for a variety of 2D colour images. First, both the proposed method and the mean-shift algorithm are independently used to filter the images. Then,

Table 1. ML-based algorithm using a two-stage mean shift

while segment borders are unclear (5-6 iterations)
  for all pixels i in image.
    Init: $\hat{\mathbf{s}}_i^0 \leftarrow \mathbf{s}_i$, $\hat{c}_i^0 \leftarrow \mathbf{c}_i$,
    while $d(\hat{\mathbf{s}}_i^{n+1}, \hat{\mathbf{s}}_i^n) > \epsilon$ **do**:
      $\hat{\mathbf{s}}_i^{n+1} \leftarrow \frac{\sum_j \mathbf{s}_j K_s(\mathbf{s}_j - \hat{\mathbf{s}}_i^n, P(\mathbf{c}_j^0))}{\sum_j K_s(\mathbf{s}_j - \hat{\mathbf{s}}_i^n, P(\mathbf{c}_j^0))}$
    endwhile
    $\hat{\mathbf{s}}_i^\infty := \hat{\mathbf{s}}_i^n$
    while $d(\hat{\mathbf{c}}_i^{n+1}, \hat{\mathbf{c}}_i^n) > \epsilon$ **do**:
      $\hat{\mathbf{c}}_i^{n+1} \leftarrow \frac{\sum_j \mathbf{c}_j K_c(\mathbf{c}_j - \hat{\mathbf{c}}_i^n, P(\hat{\mathbf{s}}_i^\infty))}{\sum_j K_c(\mathbf{c}_j - \hat{\mathbf{c}}_i^n, P(\hat{\mathbf{s}}_i^\infty))}$,
    endwhile
    $\mathbf{c}_i = \hat{\mathbf{c}}_i^n$
  endfor
endwhile

for both sets of filtered images, the same post-processing procedure as in [4] is applied to divide the images into segments. This post-processing consists of merging all pixels with colour difference below a given threshold T into regions and then merging all regions smaller than size M into larger regions. The threshold T used for merging segments are listed in Table 2. The spatial and range bandwidths used for both algorithms are $h_c = 32$, $h_s = 2$, respectively.

Performance evaluation: The results from the proposed filtering and equivalent results for the mean-shift algorithm are shown in Figure 1. From the figures one can see that the proposed method generates smoother segments than the mean-shift, but generates equally sharp boundaries at segment edges. To obtain an objective measure of the performance, both the uniformity measure U [7], and the evaluation function E [8] , which are defined by

$$U = 1 - \sum_{i=1}^{N} P_i \, \sigma_i^2 / \sigma_{\max}^2, \quad E = \sqrt{N} \sum_{i=1}^{N} \left(e_i^2 / \sqrt{N_i} \right)$$

respectively, were applied to the filtered images in the segmented regions. Here, P_i is the weighting factor for the segment i and equals $1/N$, where N is the number of segments. Further σ_i^2 is the variance of the original image in the region i, and $\sigma_{\max}^2 = \frac{(\mathcal{R}_{\max} - \mathcal{R}_{\min})^2}{2} + \frac{(\mathcal{G}_{\max} - \mathcal{G}_{\min})^2}{2} + \frac{(\mathcal{B}_{\max} - \mathcal{B}_{\min})^2}{2}$. e_i is the error between the original and filtered image in the segment i, and N_i is the number of pixels in the segment i. The uniformity measure is indicative of the homogeneity of the regions. The best performance is reached when $U = 1.0$. The evaluation function measures the distortion caused by segmentation, and a smaller E indicates better performance.

The results from these measures are shown in Tables 2 and 3. To compare the performance, the measures were calculated for two sets of tests and were applied to different images. Tables 2 shows the results from a test where the uniformity values were conditioned to be equal for both the mean-shift and the proposed method. From the results, the proposed method shows clear better performance with much smaller E values. In Tables 3, the results were obtained subject to the condition that the number of final segments would be equal for both

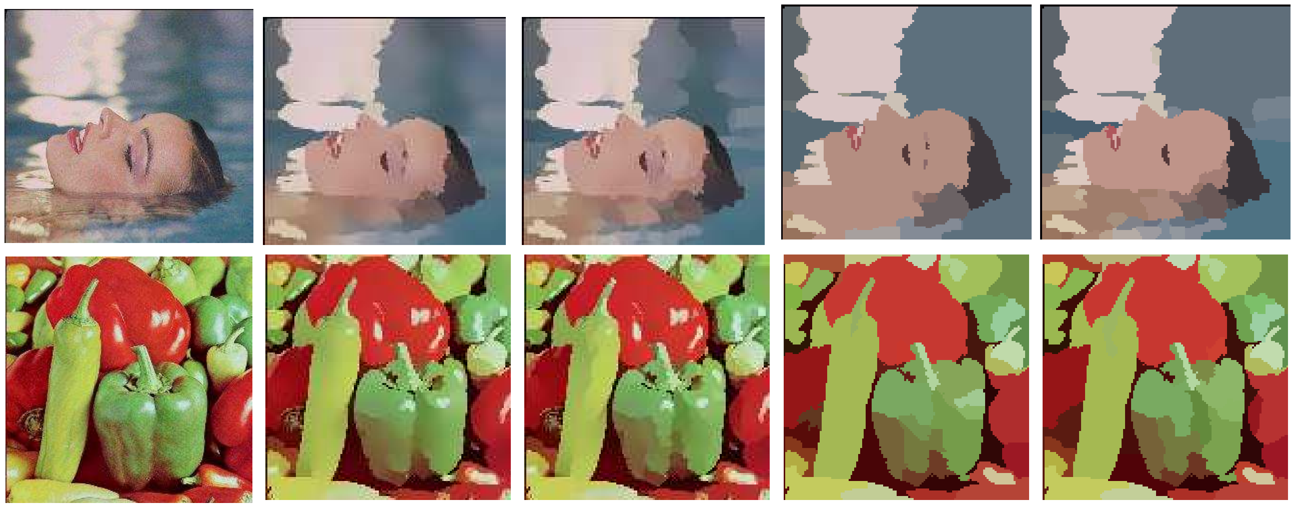

Fig. 1. Filtered and segmented images generated by both the proposed method and the standard mean shift for images: Lady and Pepper. From left to right: original, filtered image using the proposed method, filtered image using the mean-shift filter, segmented image using the proposed method and segmented using the standard mean shift.

methods. From the table, it is interesting to observe that the E values are clearly improved, however, U values are about equal or slightly decreased.

		Pepper	Monkey	Swim Lady
merging threshold t		40	40	40
E(filtered)	standard	508	1102	399
	proposed	464	683	238
U(tuned value)	standard	0.982	0.973	0.995
	proposed	0.983	0.973	0.995

Table 2. Results from performance evaluation: E and U values in (5) from the proposed method and the mean shift. In the evaluation, the uniformity for the two methods were tuned to about the same values (indicated in the bottom 2 lines).

		Pepper	Monkey	Swim Lady
merging threshold t		5	10	20
No. segments (fixed)		50	50	50
E(filtered)	standard	70	103	32
	proposed	63	94	33
E(segmented)	standard	183	178	47
	proposed	178	161	38
U(segmented)	standard	0.72	0.57	0.93
	proposed	0.73	0.57	0.92

Table 3. Results from performance evaluation: E and U values in (5) from the proposed method and the mean shift. In the evaluation, the number of segments were kept constant for both methods.

6. CONCLUSION

In this paper, an image smoothing algorithm for segmentation has been formulated and derived subject to an ML criterion. The result is related to a two-stage mean-shift algorithm which takes a new type of kernel. The method has been tested on a range of images. A visual inspection on the test results and a comparisons with the standard spatial-range mean-shift have shown a marked improvement in terms of the over-segmentation and smoothness within segments. Objective evaluations using 2 criteria functions with two different settings have shown that the proposed method has a clear improvement under one criterion function, while the performance subject to the other criterion remains almost unchanged.

7. REFERENCES

[1] S.Geman, D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images", IEEE Trans. PAMI, 6(6), 721-741, 1984.

[2] S.C.Zhu, Y.N. Wu, D.Mumford, "Filters, random fields and maximum entropy: Towards a unified theory for texture modeling", Int. Journal Computer Vision 27(2), 107-126, 1998.

[3] Y.J.Zhang (editored), Advances in Image and Video Segmentation; I.Y.H.Gu, V.Gui, "Chapter 6: Joint space-time-range mean shift-based image and video segmentation", Idea Group Inc. Publishing, 2006.

[4] D.Comaniciu, P.Meer, "Mean shift: a tobust approach towards feature space analysis", IEEE Trans. PAMI, 24(5), 603-619, 2002.

[5] C.Tomasi, R. Manduchi, "Bilateral Filtering for Gray and Color Images", in Proc. IEEE Int.Conf. ICCV, Bombay, India, 1998.

[6] S.Beucher, "Watershed, hierarchical segmentation and waterfall algorithm", in Proc. Mathematical Morphology and its Applications to Image Processing, pp.69-76, 1994.

[7] J. Liu, Y.H. Yang, Multi-resolution color image segmentation. IEEE trans. PAMI, Vol.16, pp.689-700, 1994.

[8] MD Levine, AM Nazif, "Dynamic measurement of computer generated image segmentation," IEEE Trans. PAMI, vol.7, 155-164, 1985.

[9] Barash D, Comaniciu D, "A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift," Image and Vision Computing 22 (1): 73-81 Jan 1 2004.