Local Smoothness in Terms of Variance:
The Adaptive Gaussian Filter

Giovani Gómez

ITESM Campus Morelos. Apdo. postal 99-C,

C.P. 62050. Cuernavaca, Morelos. México.

gegomez@campus.mor.itesm.mx

Abstract

Several techniques, such as adaptive smoothing [9, 10] or anisotropic diffusion [4, 5] deal with the task of local smoothing. That is, preserving principal discontinuities and smoothing within regions. Unfortunately, these types of iterative techniques have as one of their main drawbacks, the determination of the threshold on the luminance gradient. There is no way to control it easily and researchers often fall into a trial-and-error procedure. In this paper an adaptive Gaussian filter that computes directly the local amount of Gaussian smoothing in terms of variance is presented. The local variance, $\sigma^*(x, y)$, is selected, in a scale-space framework, through the *minimal description length criterion* (MDL). The MDL allows us to estimate the local smoothing in such a way that it respects the main discontinuities. The resulting smoothed image, in location x, y, is the intensity given by the convolution of the initial point $I_0(x, y)$ with its appropriate Gaussian kernel $e^{-\frac{x^2+y^2}{2\sigma^*(x,y)}}$. In fact, the proposed algorithm is not iterative, it is very stable, it is not based on derivatives nor requires any thresholds.

1 Introduction

Physical boundaries, discretised as variations of intensities, are particularly important in many low-level vision tasks, i.e. zero-crossings of second derivative, local curvature, corner and edge detection and even data compression. Detecting such discontinuities, in a real image, is not a straightforward task, because most of the algorithms are usually based on the application of derivatives, which are very sensitive to noise.

Several attempts have been proposed for dealing with noise and preserving discontinuities. Nonetheless, in most cases, the basic non-linear filters, such as median or midpoint algorithms gives poor performance. Actually, techniques such as adaptive smoothing [9, 10] and anisotropic diffusion [4, 5] provide a better way to blur within regions and not across discontinuities.

Although these last techniques offer good results, they have some drawbacks:

- The behaviour is difficult to predict.

- It is not clear how to choose the threshold on the luminance gradient.

- It is difficult to estimate the number of required iterations.

The iterative nature of adaptive smoothing or anisotropic diffusion does not allow to predict the convergence process at all, which is rarely fast.

We present an adaptive Gaussian filter that computes directly the local amount of Gaussian smoothing in terms of variance. That is by itself more intuitive than an evolution of a partial differential equation or a successive application of weighted averaging masks. The smoothing is computed in a way that respects the discontinuities. Moreover, it is not iterative, it is stable, does not require a gradient stopping function nor any threshold.

The remainder of the article is organised as follows. Section two reviews the current state of non-linear smoothing algorithms. In section three, our approach is presented and comparative examples are given in section four. Finally, in section five, we provide some concluding remarks and proposals for future work.

2   Non-linear smoothing and enhancement

Classical non-linear smoothing filters, such as median or midpoint, offer advantages of preserving more discontinuities than a simple arithmetic mean. Although these techniques are fast, in practice, they show poor results. Current techniques, such as adaptive smoothing [9, 10] and anisotropic diffusion [4, 5] deal with the task of enhancing discontinuities and smoothing within regions, at the price of increasing the computational complexity. Let us review briefly the basis of these techniques.

2.1   Scale-space techniques

The seminal work of Koenderink [2] was the introduction of the linear scale-space through the diffusion equation. In other words, the isotropic (when the diffusion process is equal for every point in the image) scale-space can be seen as the evolution of the partial differential equation $\frac{\partial I}{\partial t} = \nabla^2 I$ through the scale t. Further improvements of this idea were introduced by Perona et al. [4]. He showed that isotropic diffusion is a simplified case of the heat anisotropic diffusion equation $\frac{\partial I}{\partial t} = \nabla . c \nabla I$, where c denotes conductance, which controls the diffusion.

The blurring in anisotropic diffusion is done with the conductance c, which decreases as a function of the square of the gradient. In other words, points with "high" gradient are blurred less than points with "low" gradient. In practice that is not easy to compute, because the equation assumes that the signal is approximately piecewise constant, and is based on a model of edges as large step changes in intensity. Consequently, a single threshold on the luminance gradient can not extract the complete discontinuity, because this could vary unpredictably over the image or even along itself.

Adaptive smoothing is an iterative technique proposed by Saint-Marc et al., [9, 10]. It uses a weighted averaging mask, in spite of simple arithmetic mean, which changes as a function of local gradient. They argue in favour of repeatedly convolving the image with a small Gaussian kernel, that is, using the "cascade property". Notice that their smoothing is extremely slow and it could take an extremely large number of iterations. Moreover, the adaptive smoothing algorithm

converges at a piecewise constant image. In this way, it is completely different from its initial purpose: denoising and enhancing.

One must, however, note that the discrete approximation of anisotropic diffusion is a particular case of adaptive smoothing modifying a couple of weights in the averaging mask, see [10] for details.

Conceptually, the idea of applying successive weighted averaging masks, as a function of local gradient, is an interesting approach. Nevertheless, in practice, both algorithms presents some relevant aspects to be considered:

- The behaviour is still difficult to predict. For instance, if we choose close initial conditions (iterations and thresholds), it does not return close solutions.

- If the gradient threshold is chosen large, most of the discontinuities disappear; clearly, if the threshold is small, all discontinuities are kept, including noise.

- Their convergence process returns piecewise constant images. Those can take an extremely large number of iterations for smoothing regions. It implies, in addition, the selection of the number of iterations.

Further improvements to the original anisotropic diffusion have been developed [11, 12, 14]. Weickert uses a "tensor valued diffusion". That is, the conductance c is controlled with the direction of the eigenvectors of the second moment matrix, which are essentially first and second order derivatives. The new feature for reducing the sensitivity to noise has been too costly in terms of execution time. Whittaker et al. proposes the "vector-valued diffusion". They use multiple partial differential equations, each one computing different properties. In practice, it has been difficult to integrate it [14, 7], and it is still complex to estimate the mutual influence of coupled partial differential equations.

As a result, the edge stopping function has become more complex, and even slower to compute. In fact, new non-linear diffusion filters have not supplied better perceptual results than the original edge stopping criterion [5].

Unfortunately, adaptive smoothing and anisotropic diffusion cannot offer unsupervised smoothing, and we often fit the parameters in a trial-and-error procedure. A better way for doing unsupervised smoothing is shown in the next section.

3 Adaptive Gaussian filtering

Many existing algorithms in image processing assume a scale for feature detection, such as zero-crossings of second derivative, local curvature, corner and edge detection, to name a few. This scale represents a compromise between localisation accuracy and sensitivity to noise. Obviously, there is seldom a single scale that is appropriate for a complete image.

The local variance calculated must be useful for, at least, two basic and necessary operations in image processing: noise removal, and facilitating the localisation of features. Rewriting the last statement, we can say that the local blur, as in anisotropic diffusion or adaptive smoothing, should smooth directly within regions and not across discontinuities. More formally, in terms of Bayesian estimation,

maximising the likelihood of the observed image given the local scale, and at the same time, minimising the residual.

To see this more clearly, consider a stack of smoothed images or scale-space, $I_{\sigma_1}, I_{\sigma_2}, ..., I_{\sigma_n}$, obtained by convolving the initial image I_0 with a set of Gaussian kernels $G_\sigma, \sigma = \sigma_1, ..., \sigma_n$, where $I_\sigma = I_0 * G_\sigma$. Thus, our objective is to choose the appropriate scale (σ), from the scale-space generated, at each pair x, y. The criterion for such selection is the *optimal coding*, which must be the less complex or minimal, and, at the same time, effective for describing the parameter.

In practice, it is straightforward to estimate the *optimal code* with tools like the minimal description length principle (MDL) proposed by Rissanen [8, 1]. An image due a Gaussian filtering can be seen as:

$$I_0(x, y) = \underbrace{I_\sigma(x, y)}_{Low-Pass} + \underbrace{\varepsilon_\sigma(x, y)}_{Residual} \tag{1}$$

where the residual, called ε_σ, is the original image minus the smoothed image.

The idea of selecting the minimal description length has been successfully applied [3, 6, 15] in machine vision for balancing simplicity and accuracy. The minimal description length states that the *optimal coding* deal with the minimum number of bits that are necessary for maximising the information of both parameters. That is, the maximum smoothness with the minimal residual. Let us rewrite equation (1) as description length (dl):

$$dl_{I_0(x,y)} = dl_{I_\sigma(x,y)} + dl_{\varepsilon_\sigma(x,y)}$$

3.1 Description length of I_σ

It is not difficult to see that the measure of information, in bits, for I_0 is greater than I_σ, for $\sigma > 0$. When $\sigma \to \infty$, the measure of information over I_σ must be minimal. The reason for this fact is that the variance σ in the Gaussian filter controls the amount of information in the scale-space.

In order to calculate a fast approximation of the amount of information in I_σ, we start with the ideal model of low-pass. Allow us to consider the behaviour of signal amplitude (a) in space (x) and frequency f through the Scaling-Uncertainty Principle:

$$\mathcal{F}\{s(ax)\} = \frac{1}{a} S\left(\frac{f}{a}\right) \quad ; a \neq 0$$

Then, the Gaussian distribution in space and frequency takes the form:

$$e^{\left(-\frac{\omega^2 \sigma^2}{2}\right)} = e^{\left(-\frac{x^2}{2\sigma^2}\right)} \quad ; \sigma^2 \neq 0 \tag{2}$$

Hence, it is clear that the relation between the spatial domain ($G_{\sigma_x^2}$) is inversely proportional to the frequency domain ($G_{\sigma_\omega^2}$). The relation (2) states that $\sigma_x^2 \propto \frac{1}{\sigma_\omega^2}$. Further, the sampling theorem states that for any signal of frequency f, the number of samples s (Nyquist rate) needed for reconstructing accurately the original signal is, at least, $2f$. As we know, $f \propto \sigma_\omega^2$, that is the frequency f is proportional, given a

constant (α), to the Gaussian filter band-width, and it is controlled by its variance σ_ω^2. Thus, our sampling rate could be rewritten as $s = n(\alpha\sigma_\omega^2)$; $n \geq 2$.

Then, as a consequence of the inverse relation (2), we present the next equation, showing that the amount of samples needed for describing a smoothed image is controlled by the spatial variance of the Gaussian kernel. That is:

$$s = n\left(\frac{\alpha}{\sigma_x^2}\right) \quad ; n \geq 2$$

Although we cannot know in advance how many bits represents each s, we effectively see that it is proportional ($s \propto bits$) to the amount of information, given a constant β. Where β means the precision, in bits, used to represent I_σ. We can now summarise the above relations in the following equation:

$$dl_{I_\sigma} = n\left(\frac{\alpha\beta}{\sigma_x^2}\right) \quad ; n \geq 2 \tag{3}$$

Equation (3) estimates the description length, in bits, of a smoothed image given the spatial variance of the Gaussian filter. In addition, equation (3) shows that a correct description length of I_σ should be computed in a range. The fact that λ is unknown, does not, of course, means that is not possible to compute. In this way, if we find this range, our algorithm becomes completely stable. Having calculated the first part of equation (1) we now show the description length of the residual.

3.2 Description length of ε_σ

Even if we do not know the noise distribution, it will approximate to Gaussian distribution by the Central Limit Theorem (zero-mean Gaussian density). Thus, the probability distribution of noise can be written as:

$$Pr(\varepsilon) = e^{-\frac{\varepsilon^2}{2\sigma_\varepsilon^2}} \tag{4}$$

where σ_ε^2 is the variance of the noise, and ε^2 means the local quadratic residual[1] between the original image, I_0, and the filtered image I_σ. As can be seen in Shannon and Rissanen [8] works, the measure of information is close related to the probability. We can compute the measure of information in bits of equation (4) applying logarithm (log_2). Thus, the description length of the residual takes the form of:

$$dl_{\varepsilon_\sigma} = k\left(\frac{\varepsilon^2}{2\sigma_\varepsilon^2}\right) \tag{5}$$

where $k = \frac{1}{ln 2}$. Once a ld_{I_σ} and ld_{ε_σ} have been definited, we may, therefore, write the local description length as the sum of both terms, (3) and (5):

$$dl_{I_0} = n\left(\frac{\alpha\beta}{\sigma_x^2}\right) + k\left(\frac{\varepsilon^2}{2\sigma_\varepsilon^2}\right)$$

Grouping terms, we have:

---

[1]In a first order neighborhood.

$$dl_{I_0} = \left(\frac{\lambda}{\sigma_x^2}\right) + \varepsilon^2 \qquad (6)$$

where $\lambda = \frac{2n\sigma_\varepsilon^2 \alpha\beta}{k}$. It is positive since that $n \geq 2$ from equation (3), and the other proportional variables are also positives. It is convenient to distinguish in equation (6), that λ represents, principally, the assumed noise variance σ_ε^2 and precision used to represent I_σ. If we empirically fix the precision, we can recognise the range where λ becomes useful.

Note that in the definition of equation (6) does not appear an edge stopping function nor gradient threshold. The minimal and maximal amount of smoothness are controlled by the σ_{min} and σ_{max} in the scale space. Moreover, applying equation (6) the filter adapts its shape (σ) in the next way: if the filter is close to the discontinuities, the MDL selects a σ close to σ_{min}; otherwise, if the filter is located in a uniform structure, the MDL selects a variance close to σ_{max}. Notice that, the Gaussian shape (σ) changes as a function of MDL. The reason this works is that in the neighborhood of a discontinuity the absolute value of the residual ε^2 grows much faster with respect to σ than in a relatively uniform region of an image. As a result, the local scale will be small close to the discontinuities, and large over the uniform regions, which permits an effective noise elimination and an accurate discontinuity preservation at the same time.

3.3   Final algorithm

Once equation (6) is defined, one may use it to compute directly the local amount of smoothing in terms of variance, and, in consequence, perform a Gaussian filtering. To be specific, compute the local description length at each pair x, y through[2] $I_0, I_{\sigma_1}, I_{\sigma_2}, ..., I_{\sigma_n}$. Then, based in the MDL principle, we choose the minimal value of dl_{I_σ} at each location x, y. This $min(dl_{I_\sigma})$ returns the optimal smoothness (σ^*) in x, y. This value represents the maximum smoothing, and, at the same time, the minimum residual (see equation 6). At this step, we have the local variance ($\sigma^*(x, y)$), which is by itself, the appropriate smoothing for regions and not across discontinuities

Finally, the output smoothed image in location x, y is the intensity $I_\sigma(x, y)$ at the selected scale $\sigma^*(x, y)$.

The local variance σ of the Gaussian at x, y, allows us the possibility of making an adaptive Gaussian filter. Each point x, y is blurring as $\sigma^*(x, y)$. That is,

$$I_{\sigma(x,y)} = I_0(x, y) * e^{\left(-\frac{x^2+y^2}{2\sigma^*(x,y)}\right)}$$

4   Examples of adaptive Gaussian filtering

As a first example, we filtered the image in figure (1a). Figure (1b) shows the local scale map computed with equation (6). A value close to σ_{min} is darker. Figures (1c) and (1d) shows the final image after 50 and 80 iterations respectively

---

[2]Scales from $\sigma_{min} = 0.4$ to $\sigma_{max} = 4.0$, increment 0.1.

of a standard anisotropic diffusion algorithm[3]. Both images present noise and excessive blurring within regions, note that the adaptive version, figure (1e), does not blur beyond regions nor across discontinuities.
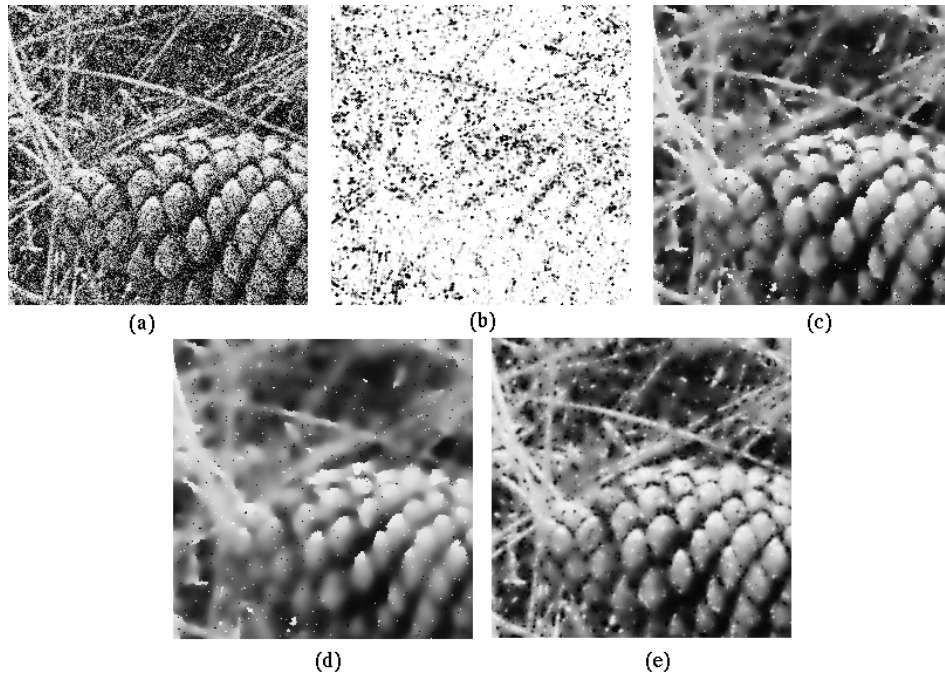

Figure 1: Example of adaptive Gaussian filtering. (a) A pine cone image corrupted with Gaussian noise, $\sigma = 40$. (b) Local scale map. (c) The output of 50 iterations of anisotropic diffusion, $k = 10$. (d) The output of 80 iterations of anisotropic diffusion with $k = 10$. (e) The adaptive Gaussian filter.

Our second example is illustrated in figure (2), where we can see the response of the filter at smooth surfaces and noise. The image of Lenna (2a) has been corrupted with Gaussian noise, and figure (2b) shows its local scale map. Figures (2c) and (2d) are computed with anisotropic diffusion, both images have noise and unwanted blurring. In fact, figure (2e) keeps its lips and shoulder, and also shows less noise. Thus, facial features are more clear.

Our third example shows a human colon image corrupted with Gaussian noise, see figure (3a). Figure (3b) shows the local scale map, $\sigma^*(x, y)$, used by the adaptive Gaussian filter. Next, figures (3c) and (3d) show clearly that anisotropic diffusion could not suppress noise properly. Figure (3e) shows an imagen with less noise and without excessive blurring.

Notice in figures (1b, 2b, 3b), that a visual discontinuity in the original images have its respective match in the local scale maps. When we compared the responses between anisotropic diffusion and the proposed algorithm, we found that it provides similar or even better results. These results could be interpreted to sup-
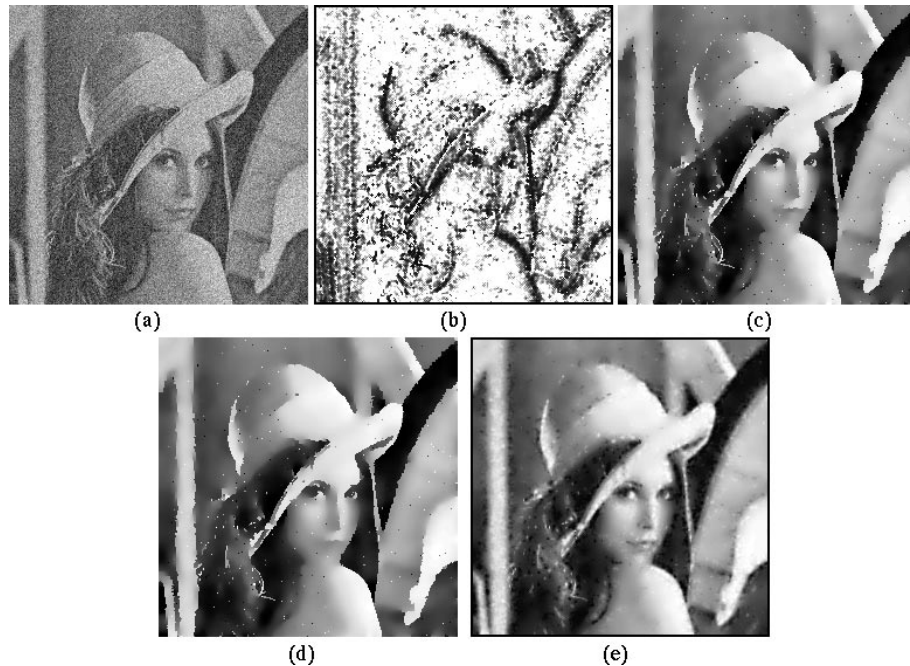
---

[3]Code available in reference [5]

Figure 2: Example of adaptive Gaussian filtering. (a) Lenna corrupted with Gaussian noise, $\sigma = 40$. (b) Local scale map. (c) The output of 60 iterations of anisotropic diffusion with $k = 10$. (d) The output of 80 iterations of anisotropic diffusion with $k = 10$. (e) The adaptive Gaussian filter.

port the idea of smoothing without edge stopping functions based on derivatives. Moreover, our experiments indicate that the output of the adaptive Gaussian filter compares favourably with the standard anisotropic diffusion algorithm, at a low computational cost, and without supplying variables.

As mentioned earlier, the algorithm is stable, i.e. not too dependent on the value of λ. For instance, λ can be within a large range with the same performance. Presumably, the range where λ becomes useful is between 2000 and 4000 (the λ is set to 3000 in all examples). The algorithm took 3 seconds per image on an IBM H50 workstation.

5 Conclusions and future work

We have presented an adaptive Gaussian filter that computes directly the local Gaussian smoothing in terms of variance. That is by itself more intuitive than a successive application of weighted averaging masks or partial differential equations. The smoothness is computed in such a way that keeps main discontinuities. In addition, our technique is not iterative, it is very stable nor requires any thresholds.

Our experiments show that the output of adaptive Gaussian filter compares favourably with anisotropic diffusion, at a low computational cost, and without
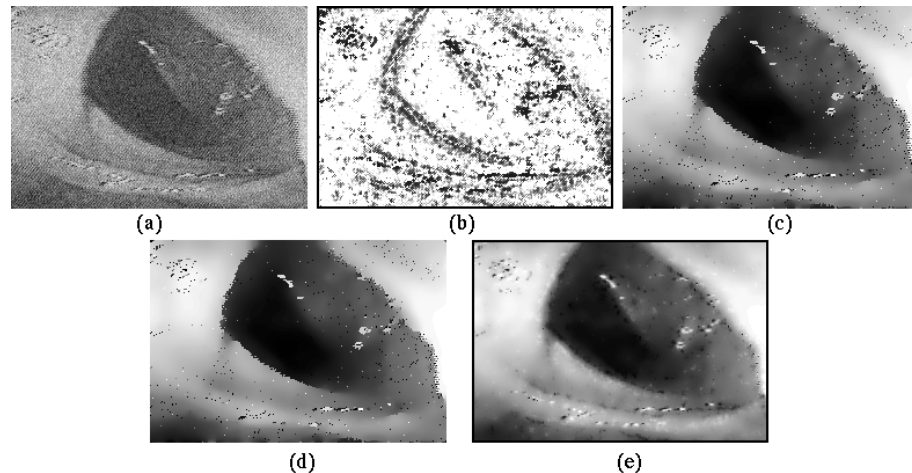
Figure 3: Example of adaptive Gaussian filtering. (a) A human colon image corrupted with Gaussian noise, $\sigma = 30$. (b) Local scale map. (c) The output of 100 iterations of anisotropic diffusion with $k = 10$. (d) The output of 120 iterations of anisotropic diffusion with $k = 10$. (e) The adaptive Gaussian filter.

supplying variables. These findings also suggest that the algorithm can contribute to the perception of different approaches, perhaps based on the MDL principle, to deal with the local smoothing. Moreover, the MDL approach shows accurate results and could be a good way to substitute the undesirable gradient threshold in other low-level techniques.

References

[1] A. Barron, J. Rissanen, B. Yu, "The minimum description length principle in coding and modeling", *IEEE Trans. on Information Theory*, vol. 44(6), pp. 2743-2760, Oct. 1998.

[2] J.J. Koenderink, "The structure of images". *Biological Cybernetics*, vol. 50, pp. 363-370, 1984.

[3] Y. Leclerc, "Constructing simple stable descriptions for image partitioning". *IJCV*, vol. 3(1), pp. 73-102, May 1989.

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

[5] P. Perona, T. Shiota, J. Malik, "Anisotropic diffusion". In B. M. ter Haar Romeny (ed.), *Geometry-driven diffusion in computer vision*, pp. 72-92. Dordrecht: Kluwer Academic Publishers, 1994.

[6] M. Pilu, R. B. Fisher, "Part segmentation from 2D edge images by the MDL criterion", *Image and Vision Computing*, vol. 15(8), pp. 563-573, Aug. 1997.

[7] M. Proesmans, E. Pauwels, L. van Gool, "Coupled geometry-driven diffusion equations for low-level vision", In B. M. ter Haar Romeny (ed.), *Geometry-driven diffusion in computer vision*, pp. 191-228. Dordrecht: Kluwer Academic Publishers, 1994.

[8] J. Rissanen, "A universal prior for integers and estimation by Minimum Description Length". *Annals of Statistics*, vol. 11(2), pp. 416-431, 1983.

[9] P. Saint-Marc, J. Chen, G. Medioni, "Adaptive smoothing: a general tool for early vision". *Proc. of the Int. Conf. on CVPR*, pp. 618-624, 1989.

[10] P. Saint-Marc, J. Chen, G. Medioni, "Adaptive smoothing: a general tool for early vision". *IEEE Trans. on PAMI*, vol. 13(6), pp. 514-529, June 1991.

[11] J. Weickert, "A review of nonlinear diffusion filtering". In B. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever (Eds.), *Scale-Space Theory in Computer Vision*, Berlin: Springer-Verlag, LNCS 1252, pp. 3-28, 1997.

[12] J. Weickert, "Coherence-enhancing diffusion filtering". *IJCV*, vol. 31, pp. 111-127, 1999.

[13] A.P. Witkin, "Scale-space filtering". *Proc. of the IJCAI*, vol. 2, pp. 1019-1022, Aug. 1983.

[14] R. Whittaker, G. Gerig, "Vector-valued diffusion". In B. M. ter Haar Romeny (ed.), *Geometry-driven diffusion in computer vision*, pp. 93-134. Dordrecht: Kluwer Academic Publishers, 1994.

[15] H. Y. Zheng, S. Blostein, "Motion-based object segmentation and estimation using the MDL principle", *IEEE Trans. on Image Processing*, vol. 4(9), pp. 1223-1235, Sep. 1995.