SEGMENTATION OF NATURAL IMAGES USING SCALE-SPACE REPRESENTATIONS: A LINEAR AND A NON-LINEAR APPROACH

Oscar Divorra Escoda, Ana Petrovic and Pierre Vandergheynst
Signal Processing Laboratory (LTS)
Swiss Federal Institute of Techonlogy in Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
WWW home page: http://ltspc4.epfl.ch
e-mail: {oscar.divorra,ana.petrovic,pierre.vandergheynst}@epfl.ch

ABSTRACT

In general purpose computer vision systems, unsupervised image analysis is mandatory in order to achieve an automatic operation. In this paper a different approach to image segmentation for natural scenes is presented. Scale-Space representation is used to extract the structure from meaningful objects in the image. Two different scale-spaces are analysed in the paper. On one hand Isotropic Diffusion (linear scale-space) is presented as the basis for an uncommitted front end, not relying on any special feature of the image. On the other hand the Total Variation Diffusion (non-linear scale-space) which makes a special emphasis on edges is also analysed. A hierarchical decomposition of the image is performed on the basis of the special characteristics of each scale-space. Iso-intensity paths will be tracked in the case of linear scale-space, whereas in the case of non-linear scale-space the evolution of level sets through scale will be tracked. In the framework of linear scale-space, the use of additional information to improve the robustness in the structure extraction is introduced. Appart from the set of several diffused versions of the image, a representation of edges through scale is included to supervise the generation of the hierarchical tree that represents the image.

1 INTRODUCTION

1.1 Scale-Space

Evidences have been found that the Human Visual System (HVS) performs some structure analysis on the incoming visual data [4, 7]. The structure of images has a close relation with multi-scale representation [4]. One of the clearest examples of multi-scale (or multi-resolution) data representation is Scale-Space [14]. Such a representation is composed by the stack of successive versions of the original data set at coarser scales. It is assumed that, the bigger the scale, the less information referred to local characteristics of the input data will appear. We also impose that general information applying to large scales will last through scale. Taking that into account, it is reasonable to think that local and high resolution scale information can be related to general and low resolution information. This will enable us to extract image structure.

1.2 Scale-Space Flavors

Scale-spaces can be generated on the basis of many different principles. It is just necessary to be able to obtain a description of image structures through scale. According to the application, it will be possible to derive the scale stack from different scale operators. In the literature, different approaches can be found. General comparisons are available in [11, 19]. A rough classification might be:

Linear Scale-Space is a one parameter family of images derived from the linear diffusion (or heat) equation [7].

Non-Linear Scale-Spaces relax the constraint of uncommitment in the processing of visual information, but keep the main properties of a scale-space [9, 10, 12, 16].

Depending on a prior knowledge about the characteristics of the images to analyse, a non-linear one can be selected. This will allow to take advantage of some special feature and will allow to preserve some image particularity.

2 LINEAR SCALE-SPACE

When there is no knowledge about the image, it is not possible to predict which will be the most advantageous scale-space. In that case, the best is to stay on the basis of an uncommitted visual front-end [17] where properties like linearity, spatial shift invariance, isotropy and scale invariance, will be kept. Such a set of properties is satisfied by the Linear Scale-Space.

Assumptions made by *Lindeberg* [17] are based on the idea of using successive convolutions to generate the scale-space. *Koenderink* first realized [7] what should be the basis for image structure analysis. Under several constraints, he defined the diffusion equation, given by (1), as the generator of its scale-space.

$$\frac{\partial I(\vec{x},t)}{\partial t} = \Delta I(\vec{x},t), \tag{1}$$

where I stands for the luminance of the image which depends on $\vec{x}$, position, and t, scale.

From (1) and from the constraint of using convolution to generate the subsequent scale levels one finds that the unique kernel that satisfies both is the Gaussian:

$$I(\vec{x}, t) = \int_D G(\vec{x} - \vec{x}', t) \cdot I(\vec{x}', 0). \qquad (2)$$

There is an important additional result. Spatial derivatives of the Gaussian are as well solutions of the diffusion equation, and together with the zeroth order Gaussian derivative they form a complete family of differential operators [17]. From this, multiscale differential analysis can thus be performed.

2.1 Edges Through Scale

The second derivative of the Gaussian is given explicitly by:

$$\nabla\mathcal{G}(x,y) = \frac{-1}{\pi\sigma^4}\left[1 - \frac{x^2+y^2}{2\sigma^2}\right]\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right). \qquad (3)$$

Instead of using it directly, we approximate eq. 3 by using the Difference of Gaussians ($\mathcal{DOG}$). To detect edges in scale, the difference between two consecutive levels of the 0th order linear scale-space is computed followed by a zero-crossing detection:

$$\mathcal{DOG}\,(x) = A_1 \exp\left(-\frac{(x-\mu)^2}{2\sigma_1^2}\right) - A_2 \exp\left(-\frac{(x-\mu)^2}{2\sigma_2^2}\right). \qquad (4)$$

where $\sigma_1 > \sigma_2$. In Fig. 1 one sees how most important edges last through scale.

The use of edge representation through scale on the basis of the second derivative of a Gaussian, is nothing else than a wavelet representation of the image. In this particular case the use of a second derivative of a Gaussian is known as the *Mexican Hat* wavelet [15]. This is another analogy with the HVS [4]. Since there are evidences of certain similarities between some parts of the HVS analysis and wavelet analysis.

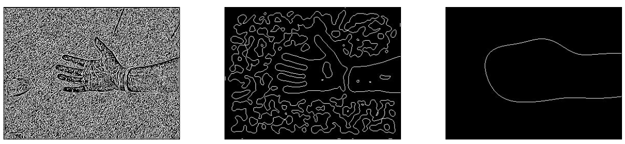

Figure 1: Edge representation through scale using DOG (levels 1,4 and 7). 1 sample per 3 octaves (first sample on the first octave)

2.2 Linear Scale-Space Segmentation: Linking up through space

The algorithm for the construction of the structure, is based on the tracking of the iso-intensity paths through scale [11]. Other algorithms where proposed relying on extrema [20, 13], but we considered to be more consistent and generic to search for the iso-intensity paths [11], since image pixels can not be fully described by extrema.

Figure 2 shows a simple schema of the idea. Levels are linked in a tree like structure. These links converge through scale according to the reduction of information imposed by the low-pass filtering.

The basic problem that arises is the search of parent pixels at a larger scale. *Vincken* [11] proposes as main linkage criteria the gray level difference between two different pixels of different neighbor levels. Those pixels having the smallest difference from a limited spatial neighborhood will be linked. That means that taking a valid pixel from a determined level (a pixel who has at least one link from the level below), a search on a circular area around that point will be performed. This search area is proportional to the inner scale.

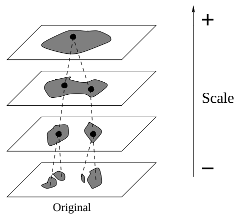

Figure 2: Hierarchical analysis of the image structure linking pixels through levels.

This linking procedure [11, 2, 1, 3] from level to level does not take into account the orientation of structures. It looks for the nearest most suitable pixel in a circular area. This is performed independently of the shape of the region where both pixels (child and parent) belong. This uncontrolled link search turns into the possibility that pixels can be linked outside the region they represent. Although it is locally true that the most similar pixels in the upper level are very likely to be the best parents for the child pixel, when search windows are large, children pixels can find sometimes better fits for their gray level some distance away from the supposed ideal pixel. In this situation, when paths evolve through scale, this small mistake turns into a divergence of a whole branch.

Figure 3 shows the algorithm we propose to reduce the divergence of paths during linking. When looking for the relation between two pixels, we test if they belong to the same region or blob at that scale. This means that when looking for linkage, all those links that cross an edge of the second derivative representation at the same scale level will not be taken into account. It follows that the area of search for a parent pixel is modified. Only that area that is included into the blob of the child pixel is taken into account in the search window.

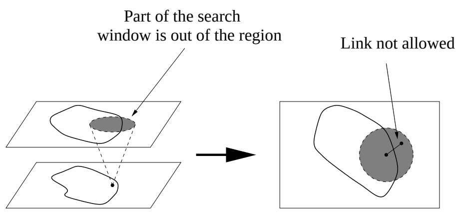

Figure 3: Wrong linkage problem.

2.3 Segmentation Experiments

2.3.1 Edge Supervision Influence

Edge detection is intended to avoid incorrect linking between different regions separated by an edge. In Fig. 4 we see the effect of the use of edges. Both segmentations are computed using the same parameters, and are segmented on the basis of the same scale level. The only difference is in the use of edges to supervise the correct linking.

Figure 4: Comparison of the effect of edge detection on the segmentation. Segmentation of the image Sergi. Level of segmentation: $\sigma_n = 25$ pixels (left: not using edges, right: using edges).

An improvement is clearly seen. In the Fig. 4 the most relevant details are signaled where the use of edges are more influent. In Fig. 4 (left) we see how part of the head is merged to the body, and next to the picture on the wall, there is a little box, which does not appear on the segmentation without edges. In Fig. 4 (right), since we use the edges at each scale, we keep from linking through them, and we success in avoiding the incorrect linking of the head, improving the definition of the contours. Finally the region that defines the box on the wall is kept, and not wrongly merged.

2.3.2 Scale Selection

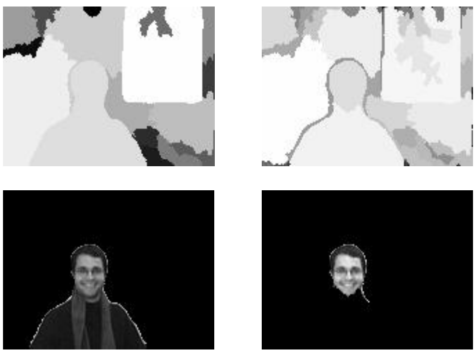

Figure 5: Obtainment of meaningful objects (down) using the Scale-Space segmentation (regions up). Level of segmentation are left: $\sigma_n = 30$ pixels, right: $\sigma_n = 25$ pixels.

Image structure gives a hierarchical description of the scene through scale. As it is explained in section 2.2, in order to obtain the segments a scale level is selected. This selection contributes to set the roots of the hierarchical tree that will represent the whole segments and to implicitly select their approximate size. In the underlying idea of the present segmentation principle, this selection of roots would be carried by a high abstraction level layer. This would interpret the structures obtained from the analysis using the scale-space.

3 NON-LINEAR SCALE-SPACE: TOTAL VARIATION DIFFUSION

As shown in the previous section with the help of numerical simulations, locally supervising image edges improves segmentation results. This is easily explained in the settings of our algorithm by the fact that we don't link pixels that belong to *different* stuctures through scales. A simple way of achieving the same task in an unsupervised manner would be to use a nonlinear scale-space in which coherent structures are preserved by the flow. According to the HVS, edges are very important primitives in natural images. We emphasize that they should be conserved in order to avoid wrong linking and this paves the way to using non-linear diffusion as a natural scale-space candidate.

First studied by Perona and Malik [8] for image processing, non-linear diffusion is realized through a general Partial Differential Equation (PDE) of the form :

$$\frac{\partial I}{\partial t} = \mathrm{div}\left(g(\|\nabla I\|^2)\right) , \tag{5}$$

where ∇I is the image gradient and g is a decreasing function. The idea is to smooth out homogeneous region as in the linear heat flow, while enhancing boundaries. Interested readers are referred to [10, 6, 9] for exhaustive reviews of all associated techniques. One such example, that we will use in the following, arises when one wants to minize the Total-Variation norm of the image [5]:

$$\|I\|_{\mathrm{TV}} = \int_{\mathbb{R}^2} \|\nabla I\| . \tag{6}$$

Gradient descent of the previous equation leads to solving :

$$\frac{\partial I}{\partial t} = \mathrm{div}\left(\frac{\nabla I}{\|\nabla I\|}\right) . \tag{7}$$

An example of such a nonlinear flow is shown in Figure 7. Let us define the isolevel sets of an image I as the sets of pixels satisfying :

$$\chi_\lambda \equiv \{x,\, I(x) = \lambda\} . \tag{8}$$

Since the TV flow is an anisotropic diffusion, it will have a tendancy to smooth regular parts of the image, while preserving its edges. The net effect of this evolution is a simplification of the isolevel sets of the image : weak edges are being eroded while strong edges will last longer which means that small uncontrasted objects will be merged into prominent structures (this can already be seen in Figure 7).

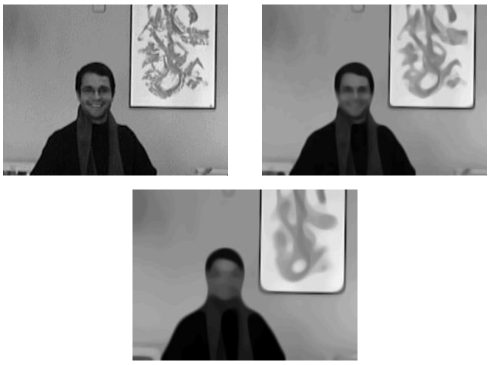

Figure 6: Example of nonlinear TV flow.

Actually it can be shown that the isolevel *curves* of the image, i.e the borders of isolevel *sets*, will move in the direction of their normal with a speed proportional to the inverse of the gradient magnitude [10].

Since edges are preserved in this new scale-space (interpreting time t as a scaling parameter), and since simplification arises at the isolevel sets stage, we can now propose a segmentation algorithm based on the ideas developed in section 2.2. We first build the nonlinear scale-space stack, $S(x,t)$, by letting the image evolve under the TV flow. $S(x,t)$ is the solution of (7) at time t with the original image as initial conditions. For each t we then compute the isolevel sets χ_λ. A thorough inspection shows that the number of these sets quickly diminishes as t increases. Moreover the edges of natural structures are being automatically handled this way. Then for two consecutive evolution times t_1 and t_2 we seek to link the corresponding isolevel sets. Let χ_{λ_1} be one of these sets at time t_1. We simply look at all the sets $\chi_{\lambda'}$ at time t_2 that overlap χ_{λ_1} and we link with the isolevel set whose level value is closer to that of χ_{λ_1}. In such a simple strategy, all level sets at time t_1 are being linked to parents at time t_2. We then manage this tree in a way similar to the linear case. The results of this algorithm are displayed in Figure 7 and show a definitive improvement with respect to the heat flow based algorithm.

Figure 7: Example of segmentation using the TV flow.

4 CONCLUSIONS

In this work we have introduced two effective segmentation algorithms. In the linear scale-space case, the use of edge supervision improves the results. On the other hand, the fact that edges are preserved in the diffusion process itself (Nonlinear case) is a great advantage of the non-linear case. Moreover, since the linking procedure is in a very preliminary stage, we consider the results very promising and very likely to be improved. In addition to the study of an optimal linking procedure, the search of other PDEs based on image processing constraints (affine and contrast invariance for example) will be very interesting.

References

[1] Niessen W. J.; Vincken K. L.;Viergever M. A. Comparison of multiscale representations for a linking-based image segmentation model. In *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, volume 21-22, pages 263–272, June 1996.

[2] Vincken K. L.; Koster A. S. E.; Viergever M. A. Probabilistic multiscale image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2), February 1997.

[3] Vincken K. L.; Niessen W. J.;Viergever M. A. Blurring strategies for image segmentation using a multiscale linking model. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR '96*, volume 18-20, pages 21–26, June 1996.

[4] Marr D. *Vision*. Freeman Publishers, 1982.

[5] L.I. Rudin; S. Osher; E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[6] Sapiro G. *Geometric Partial Differential Equations and Image Analysis*. Cambridge University Press, Cambridge, 2001.

[7] Koenderink J. J. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[8] Perona P.; Malik J. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions Patern Analysis and Machine Intelligence*, 1990.

[9] Weickert J. *Computer Vision and Applications*, chapter Design of Nonlinear Diffusion Filters. Academic Press, San Diego, 2000.

[10] Guichard F.; Moisan L.; Morel J.-M. A review of p.d.e. models in image processing and image analysis. to appear.

[11] Vincken K. *Probabilistic Multi-Scale Image Segmentation by the Hyperstack*. PhD thesis, Utrecht University, 1995.

[12] Jackway P. T.; Deriche M. Scale-space properties of multiscale morphological dilation-erosion. *IEEE Transactions on Patterb Analysis and Machine Intelligence*, 18, 1996.

[13] Lifshitz L. M.; Pizer S. M. A multi-resolution hierarchical approach to image segmentation based on intensity extrema. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(6), June 1990.

[14] Witkin A. P. Scale-space filtering. In *Proc. 8th International Conference in Artificial Intelligence*, Karlsruhe (Germany), 1983.

[15] Mallat S. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[16] Osher S.; Sethian S. Fronts propagating with curvature dependent speed: Algorithms based on the hamilton-jacobi formalism. *Computational Physics*, 1988.

[17] Lindeberg T. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[18] Lindeberg T. Scale-space: A framework for handling image structures at multiple scales. In *In Proc. CERN School of Computing*, The Netherlands, September 1996.

[19] ter Haar Romeny B. M. Introduction to scale-space theory: Multiscale geometric image analysis. Technical report, Utrecht University, 1996.

[20] Florack L. M.; ter Haar Romeny B. M.; Koenderink J. J.; Viergever M. A. Linear scale-space. *Journal of Mathematical Imaging and Vision*, 4, 1994.