Segmentation of Natural Images by Non-linear Scale-Space linking

Technical Report

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
{ana.petrovic,pierre.vandergheynst}@epfl.ch

Abstract— **A new method of image segmentation based on building up a hierarchical tree through non-linear scale-space is proposed. Non-linear scale space creates from the original image the whole family of blurred images but keeps sharp region boundaries. This characteristic of preserving edges turns out to be extremely useful in choosing a linking strategy through scale.**

Segmentation is realized in two stages. First, the set of blurred images of the scale-space are split into a number of spatial regions composed of pixels with similar intensity levels. Second, a parent-child linking between the regions of successive layers is performed based on tracking level sets through scale. Such process will generate a scale-space hierarchical tree that induces a segmentation without a priori knowledge. Experimental results based on natural images with respect to the hierarchy and segmentation are given.

Keywords— **Scale-Space, PDE , unsupervised segmentation, edge detection, hierarchical tree.**

I. Introduction

In the early stages of visual information processing, one of the most fundamental and complex tasks is segmentation. Although well posed, the segmentation problem still does not have general solution. There exists a large number of image segmentation techniques, however, most of them are based on local or global analysis. The key point in our approach is the idea of combining both of them. The segmentation technique addressed in this project is strongly related to the multiresolution paradigm that is the multiscale representation of the image. We attempt to get a global view of an image by examining it at many different scale levels.

In the recent years linear scale-space representations of images have received a lot of attention in the image processing community as it turns out to be a suitable framework for tasks like feature extraction, image restoration, noise removal, edge detection as well as segmentation. The main advantage is that this multiresolution techniques gives a deep and global knowledge of an image structure.

The idea of multi-scale analysis using the properties of the Gaussian scale space was introduced by Koenderink [1]. From this, many attempts have been carried out to build up multi-scale linking strategies, e.g by Vincken [2]. Lifshitz and Pizer [3] extended Koenderink's scheme [1] in order to build up the so called extremum stack, mainly focusing on heuristics and the performance of the algorithm. All those attempts strongly rely on the main characteristics of linear scale-space which are not promising for our approach. Our main idea is to track level sets through scale-space, hence a suitable framework for us is given by edge-preserving schemes. But linear scale-space has the inconvenient property of blurring the ”semantically meaningful” edges at coarser scale. One possibility to overcome this problem is to refer to the more general class of geometry-driven diffusion schemes preserving the region boundaries being sharp. Therefore our main idea is to use non-linear scale space in order to segment grayscale images into regions of interest (meaningful objects). In particular, we exploit the Total Variation (TV) model [4] as a special case of anisotropic diffusion.

A segmentation algorithm has been proposed for exploiting the structure of non-linear scale-space. As we blur the image, the regions bordered by edges become uniform and the smaller regions merge into larger ones at coarser scale. At a sufficiently low level of resolution, the whole image becomes one region. We attempt to construct a hierarchical tree that links all regions within the same object into a single region at a certain scale under a TV flow.

This paper is structured as follows. In section II a basic introduction to Scale-Space is given.This leads to the multi-scale hierarchy presented in section III. This section contains a detailed description of the hierarchical algorithm used for the segmentation of the image. At the end of this section, examples of several results obtained using our new technique are presented. We conclude the paper in section IV.

II. Scale-Space

Linderberg [8] observed that objects in the world appear in different ways depending on the scale of observation. He gives as a simple example the concept of a branch of tree which makes sense only at a scale from a few centimeters to at most a few meters, while it is meaningless to discuss the tree concept at the nanometer or kilometer level. Besides this multi-scale properties of real-world objects, it is necessary to cope with the complexity of unknown scenes and noise. Due to various degradation sources like the noise from the image acquisition system or the ” salt-and-paper” noise that usually occurs due to communication channel transmission errors, the information within a single image is not sufficient to obtain a reliable segmentation. Therefore, in general, segmentation methods that operate on multiple images, either connected in space or time, will improve the final segmentation. This brings us to the conclusion that for a deep understanding of the image structure, multi-resolution image representation is necessary.

One possible representation is to embed the image in a one-parameter family of images. Starting from the data $I(\vec{x})$, one creates a family of images $I(\vec{x}, t)$, where t measures the "scale" or the "time", by setting $I(\vec{x}, t = 0) = I(\vec{x})$. This method derives, from one image, a hole stack of images as it shown in Fig. 1.

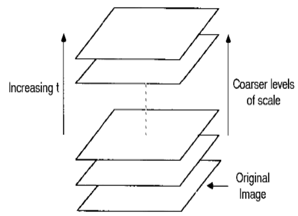


Fig. 1. Scale-Space stack.

A one-parameter family of images can be generated on the basis of many different principles. A possible derivation is the linear diffusion equation:

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

where I stands for the luminance of the image which depends on the position $\vec{x}$ and the scale t. From (1) and from the constraint of using the convolution in order to generate the subsequent scale levels one finds that the unique kernel that satisfies both is the Gaussian function:

$$G(x, y, t) = \frac{1}{2\pi t} e^{\left(-\frac{x^2+y^2}{4t}\right)}. \tag{2}$$

This idea of generating coarser resolution images by convolving the original image with the Gaussian kernel was introduced by Witkin [9] and Koenderink [1] and the resulting structure is known as *linear*, or *Gaussian* scale-space. Linear scale-space representation has been studied in deep especially by *Lindeberg* [10] and one of the most useful results is that this set of filters enables to compute the derivatives of a discrete image.

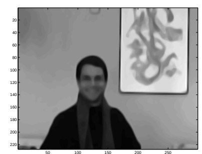
But this scale-space technique blurs important image features such as edges. In addition, it displaces the edges i.e., when moving from finer to coarser scales, it dislocates the edges. Since our main idea is to exploit intraregion smoothing through scale-space, we need an other technique that overcomes this major drawback and encourages intraregion smoothing in preference to interregion smoothing. It turns out that there exists a class of partial differential equation (PDE) based methods that lend themselves to fast numerical implementations. All of these are nonlinear models and differ in the diffusion coefficient and/or the diffusion term. A good description of these diffusion methods can be found in [11].

In this paper we used the total variation (TV) blurring and denoising method based on a variational problem with constraints using the total variation (TV) norm [4] as a nonlinear differentiable functional. The formulation of this model was introduced by Rudin and Osher [4] and the main advantage is that their solutions preserves edges very well. The detailed description of the model can be found in [12]. Here we will just give explicit numerical schemes for the 2D case:

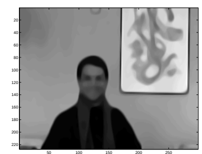
$$u_t = \frac{u_{xx}u_y^2 - 2u_{xy}u_x u_y + u_{yy}u_x^2}{u_x^2 + u_y^2}, \tag{3}$$

with $u(x, y, 0)$ being the original image. As t increases, we approach to a blurred version of our image, and the effect of the evolution should be edge detection and enhancement and smoothing as can be seen in Fig. 2. TV flow was our choice for building up a stack of blurred images because this method gives a very simple, stable and explicit procedure with the choice of parameters almost user-independent. Moreover, the implementation of this algorithm is quite simple and fast.

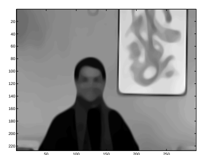
(a) Level 2



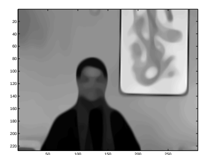
(b) Level 48



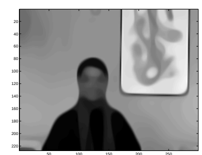
(c) Level 96



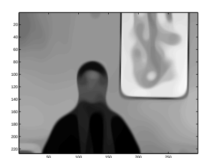
(d) Level 128



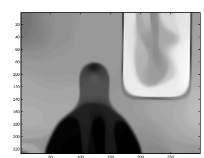
(e) Level 192



(f) Level 256



(g) Level 384



(h) Level 1024

Fig. 2. Scale-Space representation at different scale level.

Summarizing, the structure of images has a close relation with the multi-scale representation which is a TV model in our example. Such a representation is composed by the stack of successive blurred versions of the original data set at coarser scales while preserving edges. The bigger the scale, the less information referred to the local characteristics of the input data will appear, but the general information such as edges will remain through scales. 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 obtain an image segmentation scheme.

III. Segmentation Algorithm

As we already mentioned, TV flow equations perform intraregion smoothing while maintaining the edge definition. Therefore as we blur the image, smaller regions inside the object merge into larger one until it turns into a ”meaningful” object at a certain level of resolution. We want to link all those regions that become more and more uniform with respect to the scale direction. The key idea is in the constitution of a hierarchical tree under a TV flow that links all pixels within ”meaningful” object into a single point.

By partition of a given image into regions with similar intensity on each scale, we can link each region with the ”closest” one on the next scale, and build up a hierarchical tree with each of its nodes being a region. Regions at two subsequent scales can be linked only if they overlap at least in one pixel and the intensity difference is the smallest one among all the overlapping regions. Therefore, we are able to define an object as partitions of the initial image that are connected to the same regions at some fixed scale layer. In this way, the segmentation is obtained by grouping regions of the initial scale level corresponding to the same region at some fixed scale t.

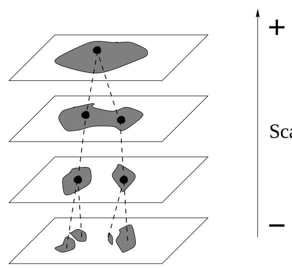


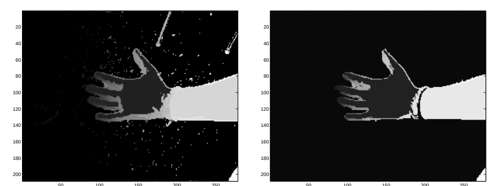
Fig. 3. Linking strategy through the scale-space.

Using the algorithm previously proposed, for any given data set, it is always possible to construct a hierarchical tree and induce a segmentation. Here is the construction procedure to implement the proposed hierarchical algorithm. The iterative scheme works as follows:

SEGMENTATION ALGORITHM
Initialization $i = 0$
1. Calculate an image I_i of scale i using equation 3.
2. Split the image of scale i into regions σ_{i_k} in such a way that each region consists of neighboring pixels with intensity difference smaller than some predefined threshold T.
3. If $i > 0$ perform linking between regions of two subsequent scales $i-1$ and i. For every region at scale $i-1$ find the corresponding one at scale i with the smallest intensity differences with constrains of overlapping at least in one pixel.
4. $i = i + 1$.
5. If $i = t$ stop otherwise go to step 1.

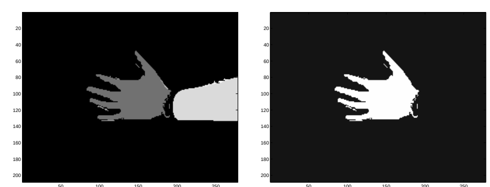
In this way, a hierarchical tree is constructed with the scale being the height. The head of each list corresponds to region on the initial scale and its tail annihilates at a certain scale level t. The image is segmented by grouping the initial regions connected to the same region at some fixed scale t. Segmentation results are presented in figures 4 and 5. We obtained these figures by applying the segmentation algorithm at different scale levels t. On the first image of Fig. 4 we can see some objects due to the noise but as we continue linking through the scale, those small objects merge into a bigger one. Obviously, as far as we go through the scale we will have the segmented image with less and less details. On the Fig. 5, we can see that if we stop our algorithm at scale $t = 512$, the hair, the face and the scarf are still recognized as an objects but at $t = 768$ the hair is already merged with the face and represents one object. Eventually, if we perform linking up to $t = 3072$, the hair, the face and the scarf are merged into one object that represents a person.

Sufficiently good results have been obtained but some work needs to be done to automatize some model parameters. Obviously, the appropriate choice of the stopping scale parameter t represents one difficulty. It can be chosen manually or if the number of objects is known we can stop at the point when we reach the predefined number of segments. Also one lack of this approach is the choice of the intralevel threshold T used for connecting neighboring pixels into a region within a single layer. Since this work is at a very preliminary stage, T was set up manually but improvements can be done for an automatic selection of the adaptive intralevel threshold.



(a) Level of segmentation: $t = 2$

(b) Level of segmentation: $t = 128$



(c) Level of segmentation: $t = 1536$

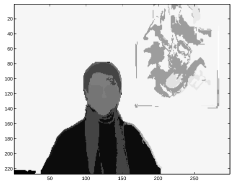
(d) Level of segmentation: $t = 4096$

Fig. 4. Segmentation results obtained applying linking strategy up to different level i of scale-space

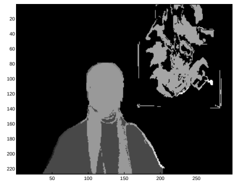
IV. Conclusions

For the integration of low-level, pixel based local image features to obtain a global object-based description we propose a fine-to-coarse scale space tracking technique. For natural images we show that non-linear scale-space tracking can be implicitly implemented by tracking level sets through scale-space.

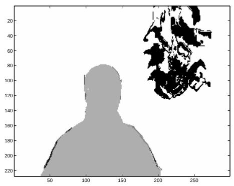
The aim of this paper is to combine the local and global analyses of the image. The proposed approach exploites the properties of non-linear scale-space taking into account many different resolution levels (global view) and gives the multi-scale linking strategy which connects neighboring pixels (local view). Two types of links are introduced to explore the deep structure of non-linear scale-space:
Interlevel links connecting two successive layers of non-linear



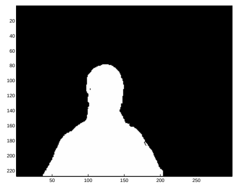
(a) Level of segmentation: $t = 512$



(b) Level of segmentation: $t = 768$



(c) Level of segmentation: $t = 3072$



(d) Level of segmentation: $t = 16384$

Fig. 5. Segmentation results obtained applying linking strategy up to different level i of scale-space

scale-space stack.
Intralevel links connecting neighboring pixels within a layer.

Therefore, useful PDE based edge-preserving methods provide a natural framework for scale-space tracking and induce segmentation without a priori knowledge.

Experimental results obtained by applying proposed algorithm are sufficiently good without any post-processing activity. Moreover, the proposed algorithm has some major advantages:

- The model is based on PDE methods that lend themselves to fast and stable numerical implementations.
- The technique gains a global view of the image by considering a multi-scale representation of it.
- The segmentation algorithm is simple in terms of implementation and computationally inexpensive.
- Since at each iteration we only work with two successive images of scale-space layers, storing the whole stack of images is avoided and just a small amount of memory is needed.

Since our research is at a very preliminary phase, we consider that the results are very promising and very likely to be improved. In addition, our future efforts will be focused on the automatic selection of model parameters and on a $3D$ implementation of our algorithm. Also, for the future work it would be interesting to apply the proposed segmentation algorithm on the other geometry-driven edge-preserving diffusions such as Beltrami Operator [5], Min/Max flow [6] and Affine Invariant Gradient Flow [7].

ACKNOWLEDGMENTS

We would like to acknowledge the contributions of Oscar Divorra Escoda. Especially we would like to thank him for his valuable comments, suggestions, corrections and ideas.

REFERENCES

[1] J.J. Koenderink, “The structure of images,” *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.
[2] A.S.E. Koster K.L.Vincken and M.A. Viergever, “Probabilistic multiscale image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 109–120, 1997.
[3] L.M. Lifshitz and S.M. Pizer, “A multiresolutional hierarchical approach to image segmentation based on intensity extrema,” *IEEE Transactions on Pattern Analysis and Machine Iteligence*, vol. 12, no. 6, pp. 529–540, June 1990.
[4] L. Rudin and S. Osher, “Total variation based image restoration with free local constraints,” in *Proc. IEEE Internat. Conf. Imag. Proc.* 1994, pp. 31–35, IEEE Press, Piscataway, NJ.
[5] R. Malladi R. Kimmel and N. Sochen, “Image processing via the beltrami operator,” in *Proceedings of Asian Conference on Computer Vision, LNC*, Hong Kong, Januar 1998, vol. 1351, pp. 574–581.
[6] R. Malladi and J.A. Sethian, “A unified approach to noise removal, image enhancement, and shape recovery,” *IEEE Transactions on Image Processing*, vol. 5, no. 11, November 1996.
[7] G. Sapiro P.J. Olver and A. Tannenbaum, *Lecture Notes in Control and Information Sciences*, vol. 219, chapter Affine invariant gradient flows, pp. 194–200, Springer-Verlag, New York, 1996.
[8] T. Lindeberg, “Scale-space: A framework for handling image structures at multiple scale,” in *Proc. CERN School of Computing Egmond aan Zee*. 8-21 September 1996, The Netherlands.
[9] A.P. Witkin, “Scale-space filtering,” in *Proceedings of the Eight International Joint Conference on Artificial Intelligence*, Karlsruhe (Germany), 1983, pp. 1019–1022.
[10] T.Linderberg, *Scale-Space Theory in Computer Vision*, The Kluwer International Series in Engineering and Comuter Science. Kluwer Academic Publishers, 1994.
[11] L. Moisan F. Guichard and J.M. Morel, “A review of p.d.e. models in image processing and image analysis,” to appear.
[12] A. Marquina and S. Osher, “Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal,” *SIAM J. on Scientific Computing*, vol. 22, no. 2, pp. 387–405, 2000.
[13] A. Marquina T. Chan and P.Mulet, “High-order total variation-based image restoration,” *SIAM J. on Scientific Computing*, vol. 22, no. 2, pp. 503–516.
[14] T. Linderberg, “Scale-space behaviour of local extrema and blobs,” *Journal of Mathematical Imaging and Vision*, vol. 1, no. 1, pp. 65–99, 1992.
[15] A. Petrovic O.D. Escoda and P. Vandergheynst, “Segmentation of natural images using scale-space representations: a linear and a non-linear approach,” Tech. Rep., EPFL, December 2001.
[16] O.D. Escoda and P. Vandergheynst, “Segmentation of natural images using scale-space representation with multi-scale edge supervised hierarchical linking,” Tech. Rep., EPFL, August 2001.
[17] A. Kuijper and L.M.J. Florack, “Hierarchical pre-segmentation without prior knowledge,” in *Proceedings of the 8th International Conference on Computer Vision 2001*, Vancouver, Canada, July 9–12 2001, pp. 487–493.
[18] R.D. Henkel, “Segmentation in scale-space,” in *Proc. of 6th Int. Conf. on Computer Analysis of Images and Pattern, CAIP’95*, Prague, 1995.
[19] B.C. Vemuri Y. Chen and L. Wang, “Image denoising and segmentation via nonlinear diffusion,” *Computers and Mathematics with Applications*, vol. 39, no. 5/6, pp. 131–149, 2000.
[20] P. Perona and J. Malik, “Scale-space and edge detections using anisotropic diffusion,” *IEEE Trans. PAMI*, vol. 12, no. 7, pp. 629–639, 1990.