

A Neural Network for Image Smoothing and Segmentation

Herbert Jahn

Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)
Institut für Weltraumsensorik
Rudower Chaussee 5
D-12484 Berlin, Germany

phone: ++49-30-67055-510
fax: ++49-30-67055-512
e-mail: herbert.jahn@dlr.de

Abstract. Older work of the author on image smoothing and segmentation with graph based parallel-sequential processing structures gave encouraging results but was not able to segment images with very strong noise. Therefore, a modified method was developed. Instead of the used “hard” Pixel Adjacency Graph (PAG) which was used formerly, a “soft” or fuzzy PAG is defined via a degree of adjacency of 4-neighbored pixels. Furthermore, the averaging over 4-neighbors is applied recursively using a nonlinear weighting function which is closely connected with the degree of adjacency and which guarantees efficient noise reduction, edge preserving, and adaptation. The discrete nonlinear dynamic equation system describing the averaging process can be implemented with a Discrete Time Cellular Neural Network (CNN). Its stable states are the smoothed images. Then the soft PAG describing the edge strength’ and the hard PAG defining the segments can be calculated. The method now can cope with strong noise. Some results demonstrate its smoothing and segmenting capability.

Keywords. image preprocessing, segmentation, edge preserving smoothing, edge detection, graphs, neural networks

1 Introduction

An attempt is made to develop a massively parallel method for image preprocessing and segmentation which can be implemented with neural networks. In the human visual system grey value segmentation as a task of pre-attentive vision is carried out by layers of neurons in or immediately after the retina [1]. Therefore, to look for a neural network approach seems to be useful.

Older work of the author [2-4] gave encouraging but not sufficient results. The method was based on a special graph, the Pixel Adjacency Graph (PAG), which can be constructed very easily by connecting 4-neighbored pixels which fulfill an adaptive adjacency criterion. The segments are understood as the connected components of the PAG. To use a graph for the description of segments seems to be adequate because segments with a complicated (fuzzy) structure can be described efficiently. Especially, this is shown by the comprehensive work on irregular, stochastic and adaptive pyramids [5-8]. Because of its usefulness the graph structure is maintained in the neural network approach presented here. This implies a cellular structure of the network

making it closely related to the Cellular Neural Networks (CNN) [9,10].

The CNN has the following structure: A neuron is assigned to each pixel of the image. Every neuron is connected with its 4-neighbors. The initial state of a neuron is given by the grey value of the corresponding pixel. In each recursion step the state of the network which represents the smoothed image is updated. The state of the network after l_{\max} recursions corresponds to the final smoothed image which can be segmented applying an adaptive adjacency criterion. That criterion which defines the PAG also determines the nonlinearity of the network via the weighting of the connections between the neurons.

The derived algorithm is closely connected to Perona-Malik anisotropic diffusion [11-13]. It should be stated too that other problems of image processing, e.g. dot pattern clustering and segmentation of simple textures, can be dealt with closely connected CNN algorithms, see Jahn [3,4,14] for a first approach.

In section 2 the neural network approach is introduced. Some results are represented in section 3.

2 Description of the method

Let's consider an image with $N \times N$ pixels (i,j) ($i,j=1,\dots,N$) and grey values $g_{i,j}$. Each pixel (i,j) corresponds to a node of the PAG. Be (i',j') a 4-neighbor of pixel (i,j) . Then the graph nodes (i,j) and (i',j') become connected by a graph edge if the adjacency criterion

$$|g_{i,j} - g_{i',j'}| \leq t_{i,j;i',j'} \quad (1)$$

is fulfilled. Applying (1) to 4-neighbors (not to 8-neighbors) guarantees the planarity of the PAG. In (1) $t_{i,j;i',j'}$ is an adaptive threshold which depends on the grey value variation within a certain neighborhood of pixels (i,j) , (i',j') (see below).

The PAG can be defined via the Neighbor Adjacency List $NAL(i,j)$ [15]. If (1) is fulfilled then (i',j') is put into $NAL(i,j)$ and (i,j) into $NAL(i',j')$. NAL defines the PAG completely. If necessary, the connected components of the PAG can be found using a graph traversal algorithm [15].

If the procedure is applied to noisy images then generally one obtains many too small segments (if the threshold t is small) or a few too large segments (if t is big). Such a behavior is not astonishing because pure local processing is used. Therefore, the idea was to smooth the grey values inside the found segments and then to apply the procedure recursively to the smoothed grey values. To formulate this in terms of nonlinear dynamic systems we consider recursive grey value averaging in the 4-neighborhood $N_4(i,j)$ of (i,j) :

$$g_{i,j}(l+1) = \frac{1}{5} \left(g_{i,j}(l) + \sum_{(i',j') \in N_4(i,j)} g_{i',j'}(l) \right). \quad (2)$$

(2) also can be written as:

$$g_{i,j}(l+1) = g_{i,j}(l) + \frac{1}{5} \sum_{(i',j') \in N_4(i,j)} [g_{i',j'}(l) - g_{i,j}(l)] \quad (3)$$

Because of its linearity, the recursive algorithm (3) with increasing recursion level l diminishes the resolution of the image and blurs the edges more and more. It is a special case of linear diffusion and can be used in the multi-scale analysis [16] of images. But here we do not want to blur edges and to smooth out image details. Therefore, the grey value differences in (3) must be weighted properly to prevent that. One can do this introducing the weights

$$w_{i,j;i',j'}(l) = \begin{cases} 1 & \text{if } |g_{i,j}(l) - g_{i',j'}(l)| < t_{i,j;i',j'}(l) \\ 0 & \text{elsewhere} \end{cases} \quad (4)$$

which represent the PAG of level l . Then (3) can be written as

$$g_{i,j}(l+1) = g_{i,j}(l) + \frac{1}{5} \sum_{(i',j') \in N_4(i,j)} w_{i,j;i',j'}(l) \cdot [g_{i',j'}(l) - g_{i,j}(l)] \quad (5)$$

(5) often gives good results but for images with strong noise the smoothing and segmentation capability of the algorithm is not good enough. To modify (5) a degree of adjacency is introduced. With $x = |g_{i,j} - g_{i',j'}| / t_{i,j;i',j'}$ the adjacency criterion (1) is equivalent to $x < 1$. Then the function

$$s_0(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & \text{if } x \geq 1 \end{cases} \quad (6)$$

can be interpreted as a degree of adjacency. More generally, each non-increasing function $s(x)$ with $s(0) = 1$ and $s(\infty) = 0$ can be defined as a degree of adjacency between 4-neighbored pixels (i,j) and (i',j') . For the computer experiments of section 3 the function

$$s(x) = \frac{1}{1+x^2} \quad (7)$$

is used. This function gives better results than the function $s_0(x)$ (6). With the function $s(x)$ the weights $w_{i,j;i',j'}$ used in (5) can be written as

$$w_{i,j;i',j'} = s\left(\frac{|g_{i,j} - g_{i',j'}|}{t_{i,j;i',j'}}\right) \quad . \quad (8)$$

Via these weights a more general definition of the Pixel Adjacency Graph can be given. Each PAG node (i,j) is connected with all four neighbor nodes $(i',j') \in N_4(i,j)$ by a PAG edge. To each graph edge between (i,j) and (i',j') the weight $w_{i,j;i',j'}$ is assigned. If the weights (8) are given by the function s_0 (6) then the PAG is a "hard" PAG with connections (weight 1) or no connections (weight 0) between the nodes. But, if another (fuzzy) adjacency degree $s(x)$ (e.g. (7)) is used then the PAG is a "soft" or fuzzy PAG.

With the weights (8) equation (5) defines a nonlinear dynamical system of equations. In principle, such a system can have stable but also unstable, periodical or even chaotic solutions. What is needed here are stable solutions for all initial conditions $g_{i,j}(0) = g_{i,j}$ ($i,j=1,\dots,N$). It can be shown that (5) is stable, because the weights which modify the stable linear algorithm (3) fulfill the condition $w \leq 1$. But convergence can be accelerated if equation (5) is modified slightly by introducing the parameter $\alpha(l)$:

$$g_{i,j}(l+1) = g_{i,j}(l) + \alpha(l) \cdot \sum_{(i',j') \in N_4(i,j)} w_{i,j;i',j'}(l) \cdot [g_{i',j'}(l) - g_{i,j}(l)] \quad . \quad (9)$$

As usual for learning algorithms stability can be enforced using a decreasing sequence of values $\alpha(l)$: $\lim_{l \rightarrow \infty} \alpha(l) = 0$. Many choices of $\alpha(l)$ are possible, but at the moment it is not clear which is best.

The quality of the method strongly depends on the choice of the threshold t in the adjacency criterion (1). The original idea for the selection of t came from dot pattern clustering [3,17] where the proximity criterion

$$d(P_i, P_j) \leq \mu \cdot \text{MIN}\{d_i, d_j\} \quad (10)$$

gave good results. In (10) $d(P_i, P_j)$ is the Euklidian distance between dots P_i and P_j , and d_i is the distance from P_i to its nearest neighbor. Therefore, it was expected that an analogous criterion with $\Delta g_{i,j}$ as the minimum of the grey value distances between pixel (i,j) and pixels (i',j') in the 4-neighborhood of (i,j) would give good results for grey value segmentation. But experiments with noisy images have shown that the adaptive threshold

$$t_{i,j;i',j'} = \mu \cdot \text{MIN}\{\Delta g_{i,j}, \Delta g_{i',j'}\} \quad (11)$$

was too busy and the noise reduction was not sufficient. Other measures of grey value

variance such as the standard deviation σ [2] have been tried with better results. But these measures have the drawback that they are influenced by edges which diminishes the separability of segments. In contrast, $\Delta g_{i,j}$ never crosses an edge. Therefore, in order to maintain that property it is useful to average the $\Delta g_{i,j}$ in a sufficient large window $W_{i,j}$ centered in (i,j) . The window size should be big enough to give stable mean values of $\Delta g_{i,j}$ (e.g. for constant grey value with additive Gaussian noise) but small enough to adapt to changing image statistics. Experiments have shown that an 11 x 11 window gives good results.

Then, for PAG construction, the symmetrical threshold

$$t_{i,j;i',j'} = \mu \cdot \text{MIN}\{\overline{\Delta g_{i,j}}, \overline{\Delta g_{i',j'}}\} \quad (12)$$

can be used. An essential property of that threshold is the *preservation of the shape* of homogeneous objects which are embedded in a homogeneous background because in such images with stepwise constant grey values there is $\overline{\Delta g_{i,j}} = 0$ for all pixels (i,j) . This is an advantage of the method over other edge preserving smoothing algorithms (e.g. the rectangular window median filter is not shape preserving).

Especially, it follows then from (1) and (12) that the very small grey value difference $|g_{i,j} - g_{i',j'}| = 1$ between regions already leads to separation. But, in 8 bit images, such small differences are not perceived. Because the results have to be assessed visually a parameter ε is introduced leading to

$$t_{i,j;i',j'} = \text{MAX}\{\mu \cdot \text{MIN}\{\overline{\Delta g_{i,j}}, \overline{\Delta g_{i',j'}}\}, \varepsilon\}. \quad (13)$$

Then, in smooth regions the threshold t is determined by ε whereas in noisy regions $\mu \cdot \text{MIN}\{\overline{\Delta g_{i,j}}, \overline{\Delta g_{i',j'}}\}$ is superior.

To use (13) for calculating the weights (8) one can compute new mean values $\overline{\Delta g_{i,j}}(l)$ in each recursion step. This gives good results but is very time consuming. Because of smoothing, it can be expected that these values decrease with increasing recursion level l . Indeed, as experiments show, it is sufficient to calculate the $\overline{\Delta g_{i,j}}(0)$ in a preprocessing step and to use the values

$$\overline{\Delta g_{i,j}}(l) = \frac{\overline{\Delta g_{i,j}}(0)}{l} \quad (14)$$

for computing the weights $w(l)$ in each iteration step.

The criterion (1) with the threshold (13) shows that the discriminability of grey value differences is diminished with increasing noise. This is in accordance with human perception. E.g. two neighbored homogeneous objects with a grey value distance of 10 each consisting of 4 pixels can be perceived effortless if the background is smooth but

this is much more difficult if there is strong noise in the vicinity of the objects.

The nonlinear difference equations (9) can be implemented with a discrete time neural network if a neuron is assigned to each pixel (i,j) . The smoothed grey value $g_{i,j}(l)$ is the state of neuron (i,j) at recursion level l , and the original image $g_{i,j}$ defines the initial state of the network ($g_{i,j}(0) = g_{i,j}$). Each neuron (i,j) is connected to its 4-neighbors in a regular grid. Therefore, the network is closely related to the discrete time version of a Cellular Neural Network originally introduced by Chua and Yang [9]. The special CNN considered here is also closely related to the time discrete version of the Perona-Malik equation [11] describing anisotropic diffusion.

After the generation of the smoothed grey values $g_{i,j}(l_{\max})$ (l_{\max} is the recursion level where stability is reached) the fuzzy or soft PAG can be generated by assigning the degree of adjacency

$$s_{i,j;i',j'} = \frac{t_{i,j;i',j'}^2}{t_{i,j;i',j'}^2 + |g_{i,j}(l_{\max}) - g_{i',j'}(l_{\max})|^2} \quad (15)$$

with

$$t_{i,j;i',j'} = \text{MAX}\{\mu \cdot \text{MIN}\{\overline{\Delta g_{i,j}(l_{\max})}, \overline{\Delta g_{i',j'}(l_{\max})}\}, \varepsilon\} \quad (16)$$

to each pair of 4-connected nodes (i,j) and (i',j') representing a graph edge. The hard PAG which is defined by the inequality (1) can be obtained from the soft PAG if the degree of adjacency (15) is binarized using the threshold $s = 1/2$. The connected components of the hard PAG are the (hard) segments of the image. These segments can be labeled with integers m , the numbers of the segments, using a graph traversal algorithm [15]. The integer function LABEL(i,j) assigns a segment number to each image pixel (i,j) .

The algorithm can also be used for edge extraction. A "hard" edge is situated between two 4-neighbored pixels (i,j) and (i',j') if the corresponding PAG nodes are not connected by a graph edge, i.e. if (1) is not fulfilled. It follows that there is always an edge if (i,j) and (i',j') belong to different segments. But there can also be (non-closed) edges inside a segment. Furthermore, via the degree of adjacency an edge strength can be assigned to every connection between 4-neighbors.

3 Results

Now some images and the results of the algorithm are presented. In all examples the recursion was stopped after l_{\max} iterations where l_{\max} was defined by

$$\text{MAX}_{i,j} |g_{i,j}(l_{\max}) - g_{i,j}(l_{\max} - 1)| < 0, 01 . \quad (17)$$

The square and background in figure 1a with a grey value distance S are superimposed with Gaussian noise: in the left part with signal-to-noise ratio $S/N=1$ and in the right part with $S/N=4$. Figure 1b shows the stable result which is obtained after 62 iterations using the parameters $\mu=3$, $\epsilon=1$. As shown by figure 1c, among the 131 segments there are two big ones representing the background and the square. Inside these big regions we have 129 small segments which represent regions generated by noise. The extreme smoothing capability of the method is demonstrated in figure 2 which displays a grey value profile across the image. The smoothing even can be enhanced if the number of iterations is increased.

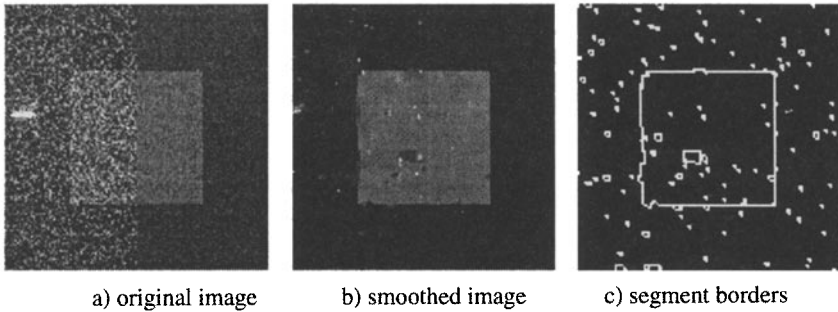


Fig. 1. Background with square and additive noise ($S/N = 1$)

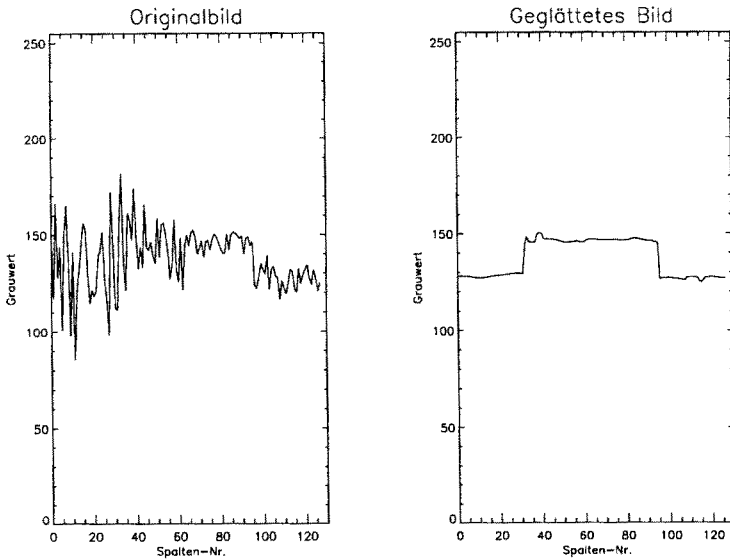


Fig. 2. Cross section through the image of fig. 1

The 128 x 128 part of the well-known Lena image (figure 3a) gives a stable result (figure 3b) with 297 segments after 76 iterations. Figure 3c shows the region borders which reveal that, at first, there are a few big segments representing parts of the face and the background and, secondly, essential detail is retained (partially) in the smaller segments.



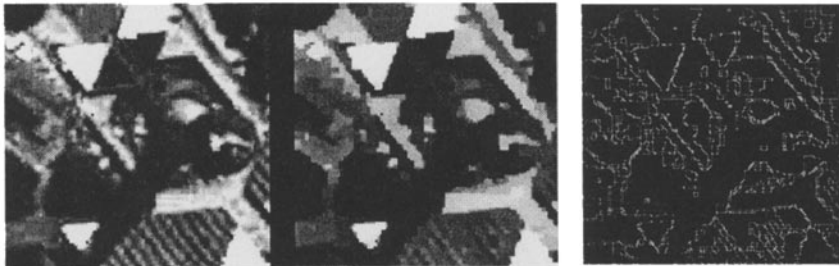
a) original image

b) smoothed image

c) segment borders

Fig. 3. Lena

Figure 4a shows the Berlin television tower imaged with the CCD Wide Angle Airborne Stereo Camera (WAAC) of the DLR from a height of 3000 m (pixel size: 1 m). In the smoothed image of figure 4b with 553 segments most of the essential detail is retained. This is shown also in the border map (figure 4c).



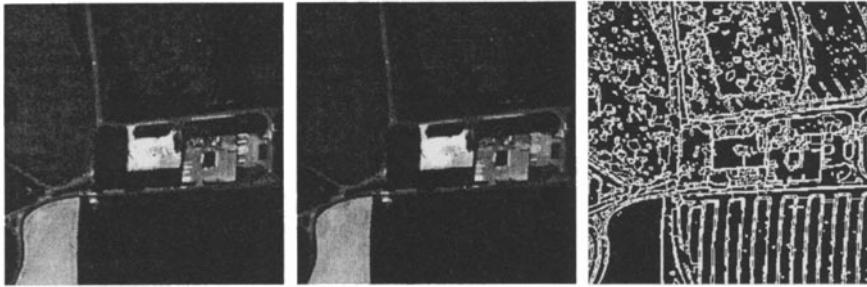
a) original image

b) smoothed image

c) segment borders

Fig. 4. Berlin television tower

The smoothed WAAC image (figure 5b), which was obtained from figure 5a with 62 iterations, was segmented after applying the grey value transform $g' = 30 \cdot \ln(g)$ taking into account Weber's law of perception. This prevents the appearance of too many small segments in the bright image regions. But, as one can see, there are also too many small segments in the dark regions. To avoid that, one can modify the logarithmic law in order to reduce the gradient of the transform $g' = f(g)$ at low g -values as proposed in [18]. The right transform should take into account the characteristics of the TV screen or printer. To fit better to visual perception detailed future study is necessary.



a) original image b) smoothed image c) segment borders

Fig. 5. Segmentation taking Weber's law into account

4 Conclusions

The few examples show that the presented neural network has a good noise cleaning capability inside big regions connected with the ability to retain essential small detail. Of course, there are also some weaknesses which easily can be seen if one compares the original images with the smoothed ones thoroughly which means that future improvements remain necessary. Especially, the possibility of reduction of the number of iterations should be investigated, the "best" adjacency degree $s(x)$ and the "best" adaptive threshold t have to be found, and, for visual assessment, the nonlinearities of the human visual system (Weber's law) and those of the displaying media must be studied. It would also be valuable to have generally accepted objective criteria for quality assessment (what is "best"?). Furthermore, the application of the method to other image features (describing texture) should be studied. Preliminary results give hope for further improvement of the method.

References

1. H.C. Nothdurft, "Common properties of visual segmentation", in: *Higher-order processing in the visual system* (Eds. G.R. Bock, J.A. Goode), J. Wiley & Sons, Chichester, 1994
2. H. Jahn, "Image segmentation with a layered graph network", SPIE Proceedings, Vol. 2662, "Nonlinear Image Processing VII", pp. 217-228, 1996
3. H. Jahn, "A graph structure for image segmentation", SPIE Proceedings, Vol. 3026, "Nonlinear Image Processing VIII", pp. 198-208, 1997
4. H. Jahn, "A graph structure for grey value and texture segmentation", GbR'97: Workshop on Graph based Representations, Lyon 1997, Computing Suppl. 12, Springer-Verlag, in press
5. J.M. Jolion, and A. Montanvert, "The Adaptive Pyramid: A Framework for 2D Image Analysis", CVGIP: Image Understanding 55, pp. 339 - 348, 1992
6. W.G. Kropatsch, and S.B. Yacoub, "A revision of pyramid segmentation", Proc. of ICPR'96, pp. 477-481, 1996
7. P. Meer, "Stochastic Image Pyramids", CVGIP 45, pp. 269 - 294, 1989
8. P. Nacken, "Image Segmentation by Connectivity Preserving Relinking in Hierarchical Graph Structures", Pattern Recognition 28, pp. 907 - 920, 1995
9. L. O. Chua and L. Yang, "Cellular neural networks: Theory", IEEE Trans. on Circuits and Systems 35, pp. 1257 - 1272, 1988

10. T. Roska and J. Vandevallé (Eds.), *Cellular Neural Networks*, J. Wiley&Sons. West Sussex, 1993
11. P. Perona, T. Shiota, and J. Malik, "Anisotropic diffusion", in: B. ter Haar Romeny (Ed.), *Geometry-Driven Diffusion in Computer Vision*, pp. 73 - 92, Kluwer Academic Publishers, Dordrecht, 1994
12. I. Bajla, "Geometry-driven diffusion - an alternative approach to image filtering/segmentation in diagnostic imaging", 6th Int. Workshop on Digital Image Processing and Computer Graphics (DIP'97), Vienna, 1997, SPIE Proceedings 3346, pp. 56-71
13. I. Bajla and I. Holländer, "Pixel dissimilarities for local controlling of the conductance in geometry-driven diffusion", 6th Int. Workshop on Digital Image Processing and Computer Graphics (DIP'97), Vienna, 1997, SPIE Proceedings 3346, pp. 120-128
14. H. Jahn, "Dot pattern clustering using a cellular neural network", 6th Int. Workshop on Digital Image Processing and Computer Graphics (DIP'97), Vienna, 1997, SPIE Proceedings 3346, pp. 298-307
15. T. Pavlidis, *Structural Pattern Recognition*, Springer-Verlag, Berlin, 1977
16. T. Lindeberg and B. M. ter Haar Romeny; "Linear Scale Space I: Basic Theory", in: B. ter Haar Romeny (Ed.), *Geometry-Driven Diffusion in Computer Vision*, pp. 1 - 38, Kluwer Academic Publishers, Dordrecht 1994
17. H. Jahn, "Eine Methode zur Clusterbildung in metrischen Räumen", *Bild & Ton* 39, pp. 362-370, 1986
18. J. Zhang and D. Wang, Image segmentation by multigrid Markov random field optimization and perceptual considerations; *J. Electronic Imaging* 7, pp. 52-60, 1998