SEGMENTATION OF REMOTE SENSING IMAGES WITH A LAYERED GRAPH NETWORK

Herbert Jahn
Head of Department
German Aerospace Research Establishment (DLR)
Institue for Space Sensor Technology
Germany

Commission III, Working Group 3

KEYWORDS: Image Analysis, Vision Networks, Segmentation, Edge Preserving Smoothing

ABSTRACT

A Layered Graph Network (LGN) for image segmentation is presented. In the LGN a graph representation of images is used. In such a Pixel Adjacency Graph (PAG) a segment is considered as a connected component. To define the PAG the layers of the network are divided into regions, and inside the regions the image is represented by sub-graphs consisting of sub-segments (nodes) which are connected by branches if they are *adjacent*. The connection of sub-segments is controlled by a special adjacency criterion which depends on the mean grey values of the sub-segments and their standard deviations. This way, the sub-segments of a layer l are merges of sub-segments of layer l-1 (the sub-segments of layer 0 are the pixels). The grey value averaging over the sub-segments is edge preserving and becomes more and more global with increasing number of the network layer. Bridge connections between the segments are prevented by the special regional structure of the network layers. The LGN can be understood as a special „neural" vision network. It is applied here to the segmentation of remote sensing images with good success.

1. INTRODUCTION

Image segmentation as one of the oldest problems in image processing and computer vision is, despite of various attempts to solve it (Haralick, 1985), not yet solved satisfactorily. Having in mind the huge capability of the human visual system, highly parallel and pipelined computation seems to be necessary for success in this field. According to (Uhr, 1980) parallel-serial layered architectures are best suited for image analysis. In this sense, a new Layered Graph Network (LGN) was developed (Jahn, 1996), which is presented and applied to the processing of remote sensing images.

Segmentation is understood here as "partial segmentation" in the sense of (Levine, 1985), i.e. the found segments do not necessarily correspond to the objects in the picture, but only to more or less homogeneous regions, which are the basis for the subsequent process of "complete segmentation", where segments correspond to objects. Crucial for this purpose is the definition of the partial segments, which according to (Pavlidis, 1977) must have a certain uniformity and which part the image into disjoint nonempty subsets. Looking at certain segments in outdoor scenes, e.g. the foliage of a tree, it becomes obvious that segments can have a complicated structure with many holes inside. Because of shading and other effects, also non-closed edges can be part of a segment.

To cope with such structures a graph representation of images (Pavlidis, 1977) seems adequate. In such a graph a segment is considered as a connected component. To define the graph each pixel must have a connection to some (0,...,4) pixels of its four-neighborhood, e.g. as a node adjacency list. Analogous to the Region Adjacency Graph (RAG) one could call such a graph a Pixel Adjacency Graph (PAG). Crucial for the construction of the graph is the criterion of adjacency of two (4-neighboured) pixels. If this criterion depends only on the gray value distribution of some local neighborhood of the pixels then, because of noise, many bridges between visually separate segments will occur, and the inherent image structure will be destroyed. Averaging over sub-segments is needed, and this can be accomplished by a special layered structure which will be presented in this paper.

The structure is similar to the well-known pyramid structure, but with the differences that, first, each layer represents a graph and averaging is carried out over the connected components of these graphs and that, second, the number of sub-segments of layer l-1 belonging to a segment of layer l is not fixed.

The criterion of adjacency of pixels or sub-segments used here depends on the standard deviation of the gray values of neighboured pixels or sub-segments. Therefore one can generate segments corresponding to image regions with slightly varying and noisy gray value distribution.

It is essential that the LGN presented here does not use any a priori information about objects or segments in the images being processed. It is not tailored to a special class of images and should be applicable to a big diversity of images. Furthermore, it is essential that no pixels (and no sub-segments in the higher layers of the network) are distinguished from the others. There are no seed points as in some other merging methods (region growing) and thus the sub-segment formation in the network layers can be highly parallelized which is necessary for efficient computation. Of course, the simulation of the network on a conventional von Neumann machine is very non-efficient and can be used only for demonstration of the ability of the

method.
In section 2 the method is explained in more detail. Results demonstrating the power of the approach are given in section 3.

2. The PAG and the LGN Structure

Let the image be represented by an image region of N x N pixels (i,j) (i,j = 0,...,N-1) with the gray values $g_{i,j}$ and $N = 2^n$. To constitute the PAG we consider each image point (i,j) and its 4-neighbours (i+1,j), (i-1,j), (i,j+1), (i,j-1). Be (i_1,j_1) one of the 4-neighbours of (i,j). Then it must be decided whether (i_1,j_1) is adjacent to (i,j) or not by using an appropriate criterion. If they are adjacent then the points (i,j) and (i_1,j_1) (which correspond to the nodes of the graph) will be connected by a branch. A suitable description of the graph is given by the node adjacency list (Pavlidis, 1977) where every node (i,j) has a list of its adjacent nodes. Now a segment of the image is defined as a connected component of the graph. Such a kind of graph definition was used in the clustering of dot patterns (Jahn, 1986), and one can interprete the segmentation method presented here as a method of clustering the data. Crucial for the graph structure to be generated is the criterion of adjacency of points (i,j) and (i_1,j_1). Node (i,j) and node (i_1,j_1) are adjacent, if their gray values fulfill the condition

$$\left|g_{i,j} - g_{i_1,j_1}\right| \le F \qquad (1)$$

Here F is some adaptive threshold which depends on the gray values in some neighbourhood of the points (i,j) and (i_1,j_1). To specify F we start with the following consideration: The visual separation of two neighboured pixels (i,j) and (i_1,j_1) is more difficult if there is a big variation of the gray values in their neighbourhood. Therefore F should be proportional to a measure of this variation. A simple measure, which gives good results, is the standard deviation σ of the gray values in a neighbourhood N of (i,j) and (i_1,j_1). To simplify the computation, it is useful to attach to each pixel (i,j) a value $\sigma_{i,j}$. It turned out that the computation of σ in the 8-neighbourhood N_8 of (i,j) is sufficient. Therefore,

$$F = t_1 \bar{\sigma}(i,j;i_1,j_1) \; . \qquad (2)$$

Here, t_1 is a threshold, and

$$\bar{\sigma}(i,j;i_1,j_1) = \frac{1}{2}(\sigma_{i,j} + \sigma_{i_1,j_1}) \; . \qquad (3)$$

The threshold (2) which increases with noise but vanishes if the grey values in the 8-neighbourhoods of (i,j) and (i_1,j_1) are constant is not sufficient.
In order to assess the segmentation results by visual inspection using a computer screen, the properties of this screen and that of the human visual system must be taken into account. Looking at such a screen one can not discriminate pixels with

$$\left|g_{i,j} - g_{i_1,j_1}\right| \le t_2 \; . \qquad (4)$$

In general, the threshold t_2 depends on the brightness (see below). Now, combining (2) and (4), the adjacency criterion for level l = 1 is

$$\left|g_{i,j} - g_{i_1,j_1}\right| \le MAX\{t_1 \bar{\sigma}(i,j;i_1,j_1), t_2\} \; . \qquad (5)$$

In principle, one could apply the adjacency criterion (5) to the whole image. Then one would obtain a graph in one step (i.e. without a layered processing structure). But this has the drawback that only local information contributes to the graph and no sufficient noise reduction takes place. Single noisy image points then can connect visually separate segments, and one generally obtains too large segments. In graph theory such connections between connected components are called bridges.
To avoid (or to minimize the number of) such bridges, averaging over adequate regions is necessary, and this can be done using a layered processing structure. In order to generate such a layered structure one divides the image i,j = 0,1,...,N-1 ($N = 2^n$) in every layer (or level) l ($l=1,...,l_{max}$) into sub-regions Reg(l,k_1,k_2) each containing $2^l \times 2^l$ image points:

Reg(l,k_1,k_2) ($k_1, k_2 = 0,...,2^{n-l}-1$): $2^l k_1 \le i \le 2^l(k_1+1)-1$, $2^l k_2 \le j \le 2^l(k_2+1)-1$

Figures 5b,c,d show the regional partition (marked by dashed lines) for l = 1,2,3 and N = 8 (n = 3).
Let's consider first the layer l=1. Applying (1) to all pixels inside a 2 x 2 region Reg(1,k_1,k_2) one can attach to every pixel (i,j) a node adjacency list NAL(1,i,j) which contains all pixels from the same region which are adjacent to the pixel (i,j). In the image plane i,j = 0,...,N-1 the node adjacency list NAL defines a graph with $N_{seg}(1)$ connected components or sub-segments. Using the NAL, the sub-segments can be labeled using a graph traversal algorithm, e.g. depth first traversal (Pavlidis, 1977). In the result of this procedure each image point (i,j) obtains a label Lab(1,i,j) (i,j = 0,...,N-1). All points (i,j) with Lab(1,i,j) = m belong to the same segment m. Now, as a feature, the mean gray value <g>(1,m) can be assigned to the sub-segment m.
Now we consider network layers (levels) l =2,3,... . Input elements are the sub-segments m = 0,1,..., N_{seg}(l-1)-1 of level l-1 with the features <g>(l-1,m). Sub-segments can be adjacent if they belong to the same region Reg(l, k_1,k_2) and if they are '4-neighbours'. Generalizing (5), the adjacency of two segments m_1 and m_2 now is defined by

$$\left|\langle g\rangle(l-1,m_1) - \langle g\rangle(l-1,m_2)\right| \le MAX\{t_1(l)\,\sigma[\langle g\rangle(l-1,m); m \in N(m_1,m_2)], t_2(l)\} \qquad (6)$$

Here the standard deviation $\sigma[\langle g\rangle(l-1,m); m \in N(m_1,m_2)]$ of the values <g>(l-1,m) is given by

$$\sigma[\langle g\rangle(l-1,m); m \in N(m_1,m_2)] = \frac{1}{2}[\sigma(l-1,m_1) + \sigma(l-1,m_2)] \; . \qquad (7)$$

σ(l-1,m) is the standard deviation of the mean gray values <g>(l-1,m') of segments $m' \in N_8(m)$, and $N_8(m)$ is the 8-neighbourhood of segment m.
Using (6), for each sub-segment m of level l-1 in a region Reg(l,k1,k2) the adjacent sub-segments which belong to the same region Reg(l,k1,k2) can be identified and stored in a node adjacency list NAL(l,m). NAL(l,m) defines a new graph of level l. Again, all connected components of this special region adjacency graph can be labeled by integers label(l,m). Using Lab(l-1,i,j) and label(l,m), the new function Lab(l,i,j) can be generated by a simple updating procedure. This process can be applied recursively from layer to layer, and in layer l=lmax the final segmentation is obtained.

Figures 5b-c display these processing steps in the three layers l=1,2,3 for the 8x8-image of fig. 1a (letter i with noise point). Adjacent pixels are connected by lines which are equivalent to the branches connecting the nodes of the graph. The regions are marked by dashed lines. Fig. 5b shows the PAG for l=1 with 25 segments. The PAG for l=2 (fig. 5c) has 11 connected components, and there are 4 segments as the final result (fig. 5d).
(6) is only an example for a possible adjacency criterion. Other criteria are permitted and even necessary. Experiments with various images have shown that the criterion (6) gives sometimes (if shading is substantial) bad results if the sub-segments are too large. Better results are obtained if inclined planes f(m,i,j)=a·i+b·j+c are fitted to the gray values inside sub-segments m of level l_{max}-1 with more than n_{min} pixels (for smaller sub-segments we use f(m,i,j)=<g>(l_{max}-1,m)). Now, sub-segments m_1, m_2 of level l_{max}-1 are adjacent, if there exist two 4-neighbours $(i_1,j_1) \in m_1$ and $(i_2,j_2) \in m_2$ with

$$\left| f(m_1, i_1, j_1) - f(m_2, i_2, j_2) \right| \leq t_2(l_{max}) \quad . \quad (8)$$

(8) is applied without a partition of layer l_{max} into regions Reg(l_{max},k_1,k_2), or, with other words, Reg(l_{max},k_1,k_2) is the whole image plane. This ensures that the final segments are not confined to squared regions.
In general, the visibility of gray value differences of neighboured segments depends on their brightness. This can be taken into account if t_2-values are used which depend on f(m_1,i_1,j_1) and f(m_2,i_2,j_2). First experiments with

$$t_2(l_{max}) = \infty, \quad (9)$$

if

$$max(f(m_1, i_1, j_1), f(m_2, i_2, j_2)) < 70$$
$$or \quad (10)$$
$$min(f(m_1, i_1, j_1), f(m_2, i_2, j_2)) > 200$$

gave encouraging results. (9) and (10) express the fact that (on a computer screen or paper) segments cannot be discriminated visually if they are both very dark or very bright.
Many experiments with various images have shown that in many cases a big number of very small segments will be generated. Some of them are important and can be seen clearly but others are not visible at all. This is caused by the up to now used adjacency criteria, which do not take into account the size of the segments. Small segments become better visible if the size increases. One can take this into account if one uses a size dependend adjacency criterion in an additional layer l_{max}+1 for small segments m with less than n_s pixels. Such a segment m will be eliminated by merging with a bigger segment m' from its 4-neighbourhood if its mean gray value <g>(l_{max},m) differs from neighboured gray values of m' less than $t_3/n_{pix}(m)$. Here, $n_{pix}(m)$ is the number of pixels of segment m and t_3 is a further threshold.
The generated PAG of level l_{max}+1 is the preliminary segmentation result which is presented here. Further layers of the LGN with other adjacency criteria taking into account not only gray value and size but also shape and spatial arrangement of sub-segments should be investigated in the future in order to segment textures and recognize objects.

For display of the segmentation results (for l=l_{max}+1) it is useful to define a function g_{mean}(i,j) with the constant value <g>(l_{max}+1,m) in every point (i,j) of a segment m. Such a g_{mean} - image shows the segmentation of the original image $g_{i,j}$ (segments have constant gray level), and it often resembles the original image very much. Then the g_{mean} - image can be used as a (edge preserving) smoothed version of the original image $g_{i,j}$. But it must be stressed that the final result is expressed by the function Lab(i,j) (=Lab(l_{max}+1,i,j)) which assigns the segment number to each pixel (i,j) of a segment. Therefore, a segment can be characterized by all of its pixels (i,j) with gray values $g_{i,j}$. Therefore, no information is lost and various segment features can be calculated which can be used in higher layers of the network for texture segmentation and object recognition.

3. Results

For demonstration of the ability of the method the LGN was simulated on a conventional serial computer (486 PC or SUN workstation) using the IDL language. A number of experiments with simulated and real world images was carried out in order to identify useful thresholds t_1(l) and t_2(l). It turned out that t_1(l) = 0.6...0.65 (l = 1,...,l_{max}-1) and t_2(l) = 5...7 (l = 1,...,l_{max}) with l_{max} = 5 gave best results. Sub-segments of level l_{max}-1 with more than n_{min} = 32 pixels were fitted by inclined planes f(m,i,j) used in (8). For the elimination of small segments with less than n_s=5 pixels t_3=30 is a good threshold value. The following results were obtained by using these parameter values.

Figure 1a shows a 128 x 128 Mars image. The image was segmented with t_1=0.6, t_2=5 (without (9), (10)). The result (fig.1b) shows 803 segments. After the elimination of small segments only 367 segments are left (fig.1c). There is no visible difference between figures 1b and 1c. Essential small details (e.g. small craters) are retained. Fig.1d shows the 3 biggest segments.

A 128 x 128 LANDSAT TM image (fig.2a) was segmented using the parameters t_1 = 0.6, t_2 = 5 (again without (9),

(10)). The result (fig.2b) has 1236 segments, the biggest segment with 2056 pixels. After the elimination of small segments we have only 744 segments (fig.2c) which are sufficient for describing this scene with many small details. The 10 biggest segments are shown in fig.2d. One can see the complicated structure of some of these segments.

The image of fig.3a (a fjord region) was segmented with t_1=0.6 and t_2=6. Without (9) and (10), after the elimination of small segments one obtains 1702 segments (fig.3b). Taking into account (9) and (10) only 1408 segments are left (fig.3c). These segments describe most of the scene adequately, but the segmentation of the clouds (lower left) is not sufficient. This shows that a modification of (10) must be investigated. The 3 biggest segments (fig.3d) once again show a complicated, 'fuzzy' structure which is best described by the PAG.

The forest image (fig.4a) which was segmented with t_1=0.65, t_2=7 and taking into account (9) and (10) has 2301 segments (fig.4b). This number reduces to 1510 after eliminating small segments (fig.4c). Many of the small segments which are retained represent the forest textures, and hopefully can be used as texture elements for texture segmentation in higher layers of the network. The 10 biggest segments (fig.4d) mainly represent parts of the background.

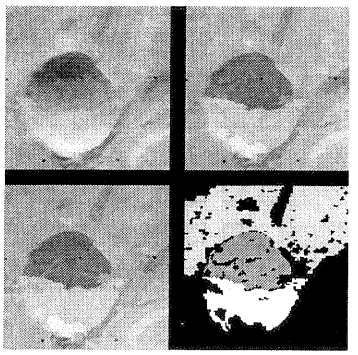

Figure 1: Mars image

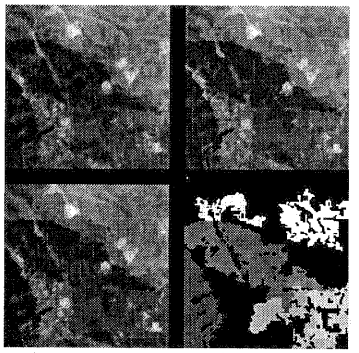

Figure 2: Landsat image (Athens)

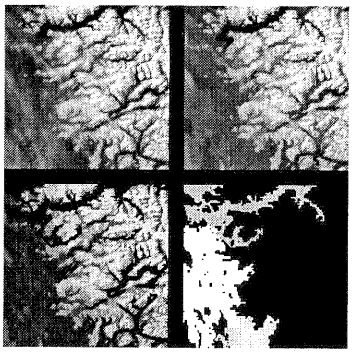

Figure 3: Fjord region

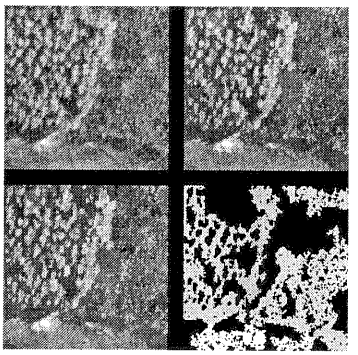

Figure 4: Forest image

References

Haralick, R.M., Shapiro, L.G., 1985. Image Segmentation Techniques. CVGIP 29, pp. 100 -132.

Jahn, H., 1986. Eine Methode zur Clusterbildung in metrischen Räumen. Bild & Ton 39, pp. 362 - 370.

Jahn, H., 1996. Image Segmentation with a Layered Graph Network. SPIE Proceedings, Vol. 2662, 1996 (in press)

Levine, M.D., 1985. Vision in Man and Machine. Mc Graw-Hill, New York.

Pavlidis, T., 1977. Structural Pattern Recognition. Springer-Verlag, Berlin.

Uhr, L., 1980. Psychological Motivation and Underlying Concepts. In: Tanimoto, S., Klinger, A. (Eds.). Structured Computer Vision. Academic Press, New York.

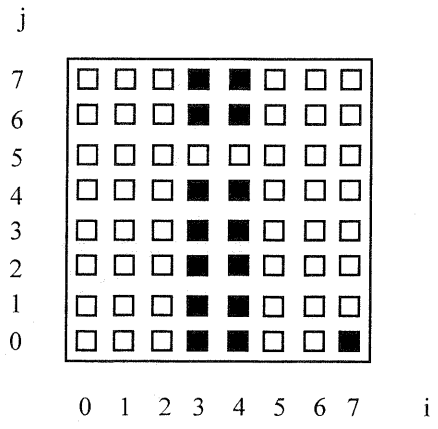

Figure 5a: Original image (layer 0)

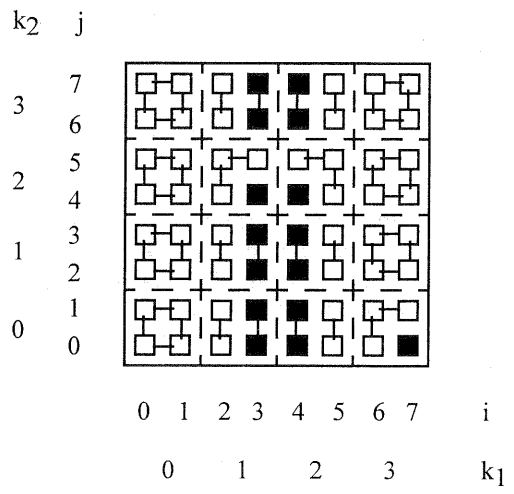

Figure 5b: Layer 1

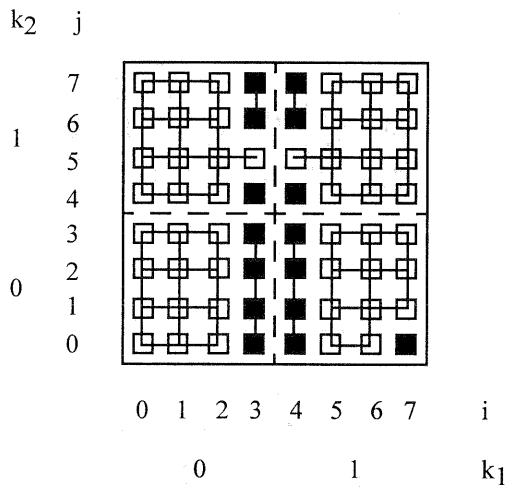

Figure 5c: Layer 2

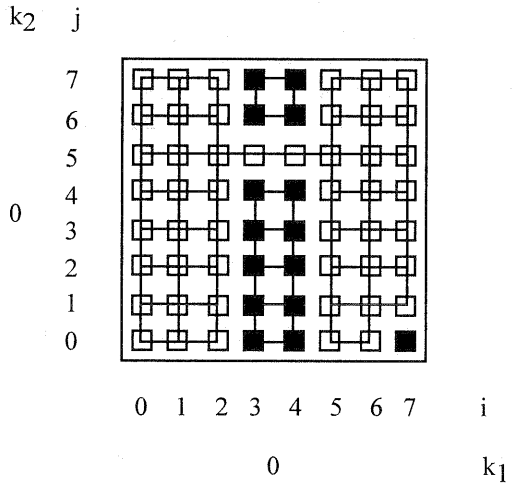

Figure 5d: Final result (layer 3)

Figure 5: PAG and regions (dashed) in the layers 1-3