RETURN

SPATIAL INFORMATION FOR IMAGE SEGMENTATION

Jorge Silva Centeno[(1)] and Manuel Weindorf[(2)]

(1) Departamento de Geomática – UFPR
Caixa Postal 19011 - Centro Politécnico
81.531.990 Curitiba - Brazil
Centeno@geoc.ufpr.br

(2) IPF - Karlsruhe University
Englerstr. 7
76128 Karlsruhe Germany
Weindorf@ipf.bau-verm.uni-karlsruhe.de

KEY WORDS : pattern recognition, spatial analysis, contextual information.

ABSTRACT

This paper deals with the vectorization of regions in the bidimensional space, defined by the image plane. The problem to be solved is the extraction of the contours after the segmentation of the pixels. Since not every region within an image can be handled as an homogeneous region, where all pixels have the same label (greyvalue), the paper concentrates on the spatial analysis of speckled pixels to group them into semantically meaningful regions. For this purpose, three approaches are used and their results are analysed and compared. The first aproach is based on the analysis of the co-occurrence of labels at neighbouring pixels. The second one performs pixel grouping based on a Delaunay triangulation and the last one resumes the regions using the watershed approach. The methods are used to segment digital images of an urban scene. The inclusion of contextual information proves to be apropriate for the segmentation of hatched regions and to resume spreaded elements. The results are compared and discussed, showing the positive and negative aspects of each approach.

1 INTRODUCTION

Graphic recognition has developed for computer assisted analysis of digital images. Segmentation and vectorization are a special track in this field, because they enable a representation on which image understanding algorithms operate (Doermann, 1998).These basic operations are currently applied to a much larger domain than was originally intended, since they are used as an intermediate step in the process of converting original data to a more expressive representation (Kunz et al., 1997). For the documentation of architecture and cultural heritage in cities, the availability of digital photography has proved a great potential. In urban scenes, most of the elements, like walls, can be segmented according to their spectral properties, since they are more or less plain surfaces with an homogeneous colour. Nevertheless, walls also contain other significant architectural or artistical elements, which can not be identified considering the wall as an homogeneous surface. Therefore, approaches which allow the identification and the grouping of these elements are needed. A pure spectral analysis is not sufficient to perform this task. In the following sections, we discuss three alternatives for including spatial information in the analysis.

2 SPATIAL FEATURES

In a very simple image, all regions would appear homogeneous and easily separable from the neighbourhood. This is not very common in urban scenes, since many portions of digital images present a speckle effect or some small elements appear spreaded over the scene. The set of spreaded elements may form a bidimensional region, which is characterized by its structure. The task of low-level image processing is to separate this regions from the rest of the image and estimate the corresponding contours.

As Pratt states (Pratt, 1974), textured regions can be classified as being artificial or natural. Artificial textures are produced by the arrangment of symbols placed against a neutral background. The symbols may be line segments, dots, characters or small figures. An example is shown in fig. 1 , where the repetition of small figures on a wall characterizes a region in a bidimensional space. Natural textures occur at natural scenes and consist of a semi-repetitive arrangement of pixels. It is produced when the objects in the scene are smaller than the image spatial resolution. Examples are brick walls or sand in photographs.

In natural scenes of buildings, facades may also present a textured pattern, since different materials are used for its construction. The structure of the regions, in terms of spatial distribution of the gray

RETURN

levels, may give a significant amount of information for the interpretation of the material that was used and therefore for the characterization of the objects in the scene. Repetition of architectural or artistical elements may also characterize a region and, in this case, it becomes important to identify and group these elements. This can be seen at panels, like the one showed in figure 1 , where the artist used the repetition of small graphical elements to cover parts of the wall.

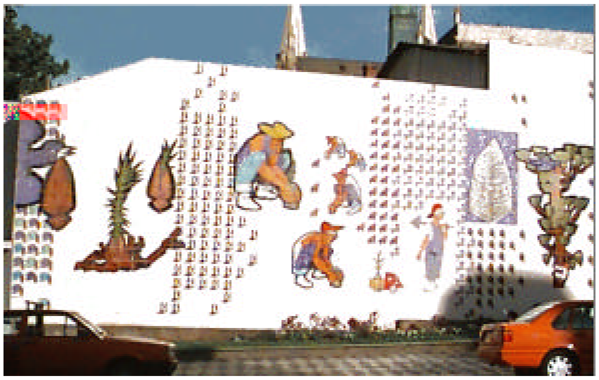

Figure 1 Digital image of a panel in Curitiba

The work showed in figure 1 belongs to Potí, a brasilian artist who painted various panels on the streets of the city of Curitiba. His panels give the city a special characteristic and identity, since they display scenes of the life in this city and its surrounding region.

3 PRE-PROCESSING

The aim of our analysis is to identify and group the different elements present on the panels like the one displayed on figure 1. For this purpose, we divide data processing in two steps: colour segmentation, which we call pre-processing, and spatial analysis.

Known colour properties of the elements allow to perform spectral segmentation of the image. The set of pixels of a colour class can be displayed as a binary image. Binarization can be performed by a multispectral classification of the image (Weindorf, 1994) or an image segmentation. Depending on the segmentation method, some pixels which should belong to a region of the same semantics may not be included and some background pixels may be labeled as false positives. This can be seen on figure 2, where some of the small elements are not complete and the wall behind the panel appears as a speckled region.

The result of this spectral image analysis are the separated colour layers. Figure 2 displays the layer of the white pixels of the image displayed on figure 1. We will use this in the discussion that follows.

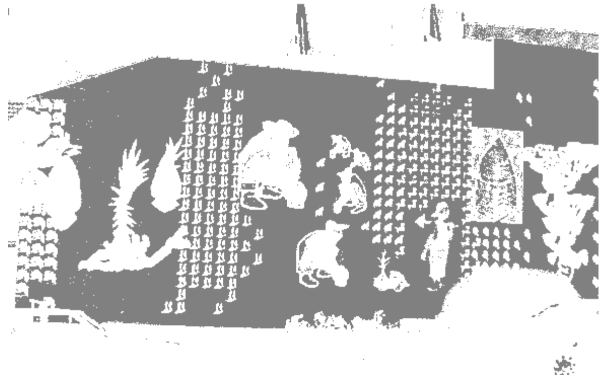

Figure 2: Binary image after the segmentation of the white pixels.

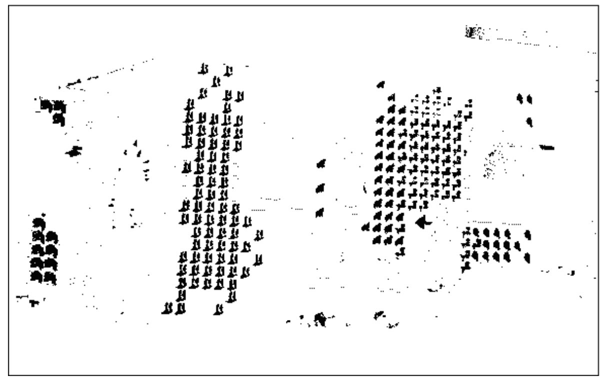

Figure 3: Small groups of pixels

After the extraction of the larger figures, the smaller figures, together with some noise, remain and can be stored in a new binary layer (figure 3). This step of the process is the starting point for the spatial analysis.

4 SPATIAL ANALYSIS

We base the analysis on a definition of textured regions presented by Hawkins (Hawkins, 1970). The textured regions we consider are regions where a local 'order' is repeated.This order consists in an arrangement of elementary parts, which are in turn roughly uniform entitites having approximately the same dimensions everywhere within the region (Hawkins, 1970). This definition is very broad and since the borders of such regions are not explicitily defined, it doesn't lead to a simple measure for the quantitative definition of edges (Pratt, 1974). Nevertheless, it helps us to consider two important components for this kind of regions: the labeled pixels, which may form symbols, and the inner spaces between them, which allow to perceive the spatial arrangement of the elements.

We focus our attention on the inner spaces. Despite errors, the background spaces inbetween the labeled pixels can be considered as inner regions, since they are surrounded by labeled pixels.

Here we understand an *inner* region as a region that is located within the bi-dimensional space defined by at least three labeled pixels. Considering three non-collinear labeled pixels pixels q_a, q_b and q_c, their inner common space is defined by the set of points x that satisfy:

$$x = \boldsymbol{a}(\overline{q_a, q_b}) + \boldsymbol{b}(\overline{q_a, q_c}) \quad (1)$$

$$0.0 < \boldsymbol{a} + \boldsymbol{b} < 1.0 \quad (2)$$

The identification of the inner spaces of a region and its classification in terms of their geometry (area or radius) leads to the segmentation of the image in terms of texture. For instance, an homogeneous region would not show inner spaces. On the other hand, a textured region, as the one displayed on figure 3, is characterized by regular inner spaces. Therefore, the task consists in identifying this spaces and their adjacent labeled pixels, and grouping them to form larger regions.

4.1 Coocurrence matrix

One valid appproach to segment textured regions is to compute a measure of coarseness or homogenity within the neighbourhood of each pixel in the image. The spatial correlation between pixels has been suggested as the base of a texture measure (Pratt, 1974). The neighbourhood of each pixel is analysed and a representative value is stored for each point of the image. In this manner, a second image is generated from the original one. Its pixels store an estimated value of the variation of the grey levels within the specified neighbourhood.

A set of texture features is described in in (Haralick and Shapiro, 1997). This approach defines a matrix of relative frequencies describing the appearance of two grey levels separated by a predefined distance and angle. From this co-occurrence matrix, features as local homogenity, contrast, correlation and entropy can be computed for each pixel.

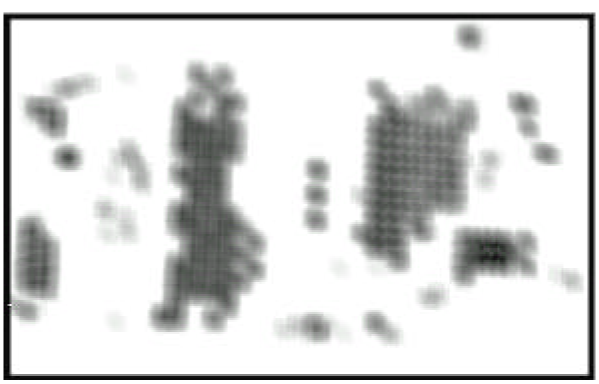

Figure 4: Local homogenity

Figure 4 provides an example of an image produced by computing the local homogenity of the image showed on figure 3. Dark pixels of figure 4 are associated to low values of local homogenity. This image can be binarized by choosing a maximum value, separating pixels with lower local homogenity. Since they are located close to the original labeled pixels, they give an idea of the inner spaces between labeled pixels.

4.2 Triangulation

In order to combine structured image regions into semantically homogeneous clusters, it is necessary to use mid-level image processing procedures going beyond grayvalue based methods.

In the presented approach the analysis is based on the relational properties of the segmented features represented internally in a feature graph structure based on the delaunay triangulation. In this graph the nodes are represented by the Delaunay triangles and the edges are represented by the neighbourhood of the triangles.

The segmentation process based on the triangulation uses Delaunay triangles as basic primitives instead of the spectral information of the pixels. This means after the image is segmented by low-level image analysis a Delaunay triangulation (Delaunay tesselation) has to be built up to serve as input of the following interpretation process. Tesselations of space reflect the spatial relationship between points, since they are composed of line segments joining neighbouring elements. For the purpose of the present study we refer to the Delaunay triangulation computed from the labeled pixels in the bidimensional space of the image.A triangulation is a subdivision of an area into triangles. The Delaunay triangulation, a special case of triangulation, has the property that the circumcircles of every triangle are empty circumcircles (Okabe et al., 1992).

Given a set of points P in a subdomain Ω^2 of the bidimensional space of the image (R^2), two points p_i and p_j are connected by an edge of the triangulation if and only if there is another point $p_k \in \Omega^2$ that is equidistant to p_i and p_j and closer to p_i and p_j than to any other point $p_x \in \Omega^2$.

$$\begin{array}{lcl} edge(p_i, p_j) & exists & \Leftrightarrow \exists k \mid k \in \Omega^2 \wedge \\ d(p_i, p_k) & = & d(p_k, p_j) \wedge \\ d(p_k, p_i) & < & d(p_k, p_x) \\ \forall x \neq i, j & & \end{array} \quad (3)$$

where $d(p_i, p_i)$ stands for the distance between two points p_i and p_j.

A Delaunay triangle $D(p_i, p_j, p_k)$ exists if and only if the three edges exist and the triangle is a Delaunay triangle if and only if there is no other point inside its circumcircle. If c is the centre of the circumcircle of the triangle $D(p_i, p_j, p_k)$ and r its radius:

$D(p_i, p_j, p_k)$ *exists* « $c ½c ÎΩ² Ù (4)

$r = d(c, p_i) = d(c, p_j) = d(c, p_k)$ Ù

$r < d(c, p_x)$ "x ¹ i, j, k

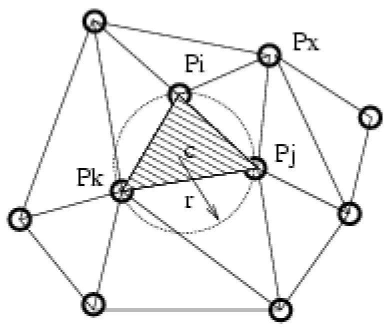

Figure 5: Delaunay triangle

As a model for the proximity (the degree of neighbourhood or the extent of the inner spaces) of three labeled pixels in the segmented image, several measures can be applied:

- Triangle perimeter p (absolute measure)
- Triangle area a (absolute measure)
- Shapefactor $sf = p^2/a$ (relative measure)

Applying thresholds for this parameters the elements of the initial Delaunay triangulation (Triangles) are labeled as *valid* or *invalid* resulting in a set of clustered triangles forming the regions (see fig. 6). The basic idea is to eliminate those triangles with larger perimeter, associated to distanced pixels, and to group the remaining triangles into significant regions of neighbouring pixels. Different values of the perimeter threshold allow the separation of regions with different densities (Weindorf, 1994), because the perimeter of the triangles is proportional to the distance of the points. In fact, an extremly low threshold value groups just adjacent pixels. On the other hand, larger values allow pixels that are not necessarily adjacent to be grouped.

To avoid the formation of islands, groups of triangles marked as invalid can be reactivated by applying an area criterion. To identify this enclosed islands, a depth-first search in the graph is used, starting at a triangle marked as invalid. In the end valid and adjacent triangles are melted together and form closed polygons in vector format. The result is a set of polygons describing the regions of the image with unique structure.

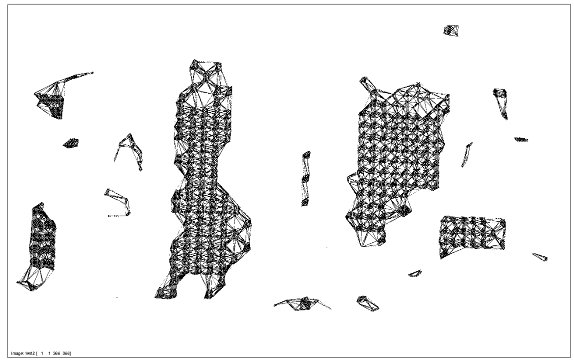

Figure 6: Triangle clusters

4.3 Watershed approach

An approach similar to the one used to delineate watersheds from digital elevation models, as described by (Jenson and Domingue, 1988); (Mark, 1983), can be used to identify the inner spaces.The main idea of such algorithms consists of simulating flow over a surface, the elevation matrix, and defining preferential flow paths based on the local gradient. When dealing with raster DTMs, the flow can occur from a pixel to one of its adjacent neighbours. At each pixel within the DTM, the local gradient defines the preferential flow direction and the set of directions over the matrix defines flow paths, which have the constraint that they have to point only downhill. If $h(x)$ is the altitude at pixel x, a flow direction $e(x,v)$ towards the neighbouring pixel v associated with a larger local gradient ($g(x,v)$) is assigned to x.

$$e(x) = \{e(x,v) \mid g(x,v)\} = \max(g(x,i))\} \qquad (5)$$

$$g(x,i) = \frac{h(x) - h(i)}{d(x,i)} \geq 0.0 \qquad (6)$$

Since a DTM may have depressions where an uphill flow would be necessary, it is practical to identify and fill them before estimating the flow paths. This task is performed iteratively, marking pixels with undefined flow direction and filling their watershed up to a value of $h(x)$ that satisfies the imposed restrictions.

Although this algorithm was originally developed to delineate watersheds in DTM's, it has also been extended to digital image processing. Examples of the application of the watershed segmentation to digital images can be found in (Vincent and Soile, 1991) and (Silva Centeno, 1999).

In a binary image there is no smooth variation of the pixel value similar to the elevation in a DEM, however it is possible to generate one as a function of the distance to the labeled pixels. Being d_{xy}, the distance between a background pixel x and a labeled pixel y:

$$h(x) = max(f_{xy}) \quad (7)$$

$$f_{xy} = 255 - d(x,y) \quad if \quad d_{xy}$$

$$f_{xy} = 0 \quad if \quad d_{xy} > d_m$$

Since the inner spaces between labeled pixels are relatively small, larger flowpaths can be discarded by the selection of an appropriate value for d_m. Those pixels that constitute part of a flow path that ends at a pixel with null elevation are discarded. The remaining regions are the catchments of the inner spaces that form depressions, or valleys with no outlet. This particular situation occurs at inner spaces between labeled pixels, because of the method used to generate the DTM.

The value of d_m controls the form of the DTM and also the size of the depressions, which are used to identify the inner spaces. It plays the role of a segmentation parameter, which describes a special geometrical feature of the regions, since it is responsible for the occurrence of valleys. If d_m is too small then only thin regions form depressions and small inner spaces are identified. The segmented inner regions, together with the adjacent labeled pixels form a solid region characterized by the spatial dimension of the inner spaces.

5 EXPERIMENTS

The three approaches were used to segment natural scenes. An example is shown on figure 1. This image was taken using a KODAK DC40 digital camera, that provides 756 x 504 pixel resolution, and captures 24-bit color pictures.

The image was segmented based only on its spectral properties and binary layers were obtained. The elements of interest within the wall were segmented, classified according to their size. The smaller figures and other small concentrations of pixels were stored in a binary image, as displayed on figure 3. This image was used to perform the spatial analysis using the three approaches.

The results were compared in order to evaluate the ability of each method to group the small figures and produce polygons enclosing them. Three sets of polygons were obtained, one obtained using a threshold for the local homogenity (figure 7), another using the watershed approach (figure 8) and the last one was derived from the Delaunay triangulation (figure 9).

The contours of the regions are directly available using the Delaunay triangulation, since the analysis is performed based on the triangles (resp. their edges) of the triangulation, which are already in vector format. Since the other two methods process data in raster format, contour tracing (Niemann, 1974). is necessary to obtain the edges. The polygons displayed in figure 7 were estimated from the binary image that was obtained thresholding the local homogenity.

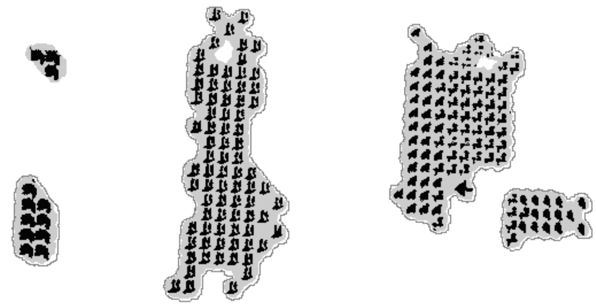

Figure 7: Segmentation using the local homogenity.

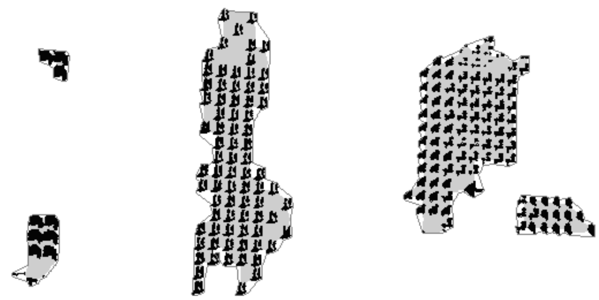

Figure 8: Segmentation using the watershed approach.

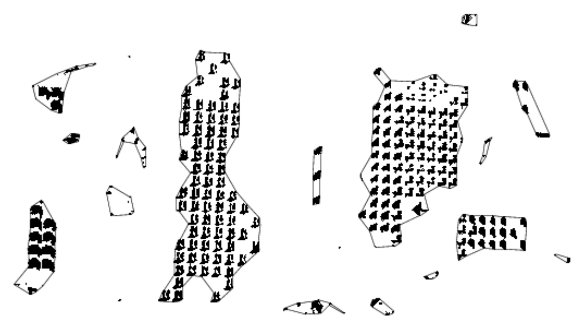

Figure 9: Segmentation using the Delaunay triangulation.

The concept of inner valleys used in the watershed approach has the drawback that it is not able to include the open spaces between labeled pixels at the borders of the region. Although the result is correct, from the point of view of inner regions, it is not satisfying because it underestimates the area covered by the ensemble (grey pixels on figure 8). Better results can be achieved by computing the hull of the polygon obtained joining the identified inner regions and the adjacent labeled pixels, as displayed on figure 10.

RETURN

Figure 10: Example of the hull of a segmented region. Labeled pixels appear in black, inner regions in grey.

6 DISCUSSION

Figures 7, 8 and 9 will be used to illustrate a comparison of the results obtained using each approach. The set of polygons obtained computing the local homogenity over the image, overestimate the total area. The choice of the size of the window and the geometry between neighbours (distance and angle) used to compute the coocurrence matrix, play an important role in the computation of such texture features and the final results. Small values produce holes within the region and larger values overestimate the area. In order to represent the texture of areas depicted using the repetition of patterns, the distance parameter has to be large enough to include two neighbouring symbols. Since lower values of the local homogenity are not restricted to the pixels within the common inner region and also occur outside its border (as seen on figure 4), it is difficult to obtain a balance that leads to a satisfying contourline. Therefore, the distance parameters influence the value of the texture features not only within the textured regions, but also at their imediate neighbourhood. The consequence is the overestimation of the area, specially for regions with large distance between symbols.

The estimation of the inner spaces using the watershed approach does not have the same problem. On the other hand, it underestimates the area, because it is based on a very hard concept of inner space, which fails especially on the borders of the region. This fact can be explained by the form of the watersheds obtained based on the pixels near the borders. Since these basins do not form depressions, that means that the simulated flow finds an outlet in the background, they are not marked. This drawback is partially compensated by computing the hull of the resulting polygon, but since some pixels and small groups of pixels are lost in the segmentation, the resulting region may be incomplete. That happens especially when the regions form peaks, as seen on figure 8. Nevertheless, this method has the advantage of performing the analysis raster based. This method does not depend on the previous definition of the direction in which the neighbours are to be searched, as the texture parameters and the only parameter to be chosen is the distance used to simulate a DTM. It also does not depend on the form of the neighbourhood, since the distance used to generate the DTM is computed in all directions.

The best result is obtained using the triangulation. This approach does not depend on a previous definition of the neighbourhood and processes data in vector format. The use of the triangle perimeters of the Delaunay triangulation as a measure of the distance between pixels shows good results. In our example regions represented by the small figures on the panel are easy to separate using this distance criterion.

The computation of the triangulation is a very complex process and demands more computational effort than the other methods. Nevertheles, once computed, it can be used to produce results using different perimeter parameters. For example, the textured wall located behind the panel on figure 3 was eliminated using a very low value of the perimeter, which allowed only very close pixels to be grouped. A larger value was used to group more distant elements, like the small set in the middle of the image (figure 9). This method also has the advantage off avoiding the estimation of the contours at the end of the process. Further, the resulting contour describes the objects of the original segmentation very precisely.

Problems occur if there are a lot of pixels labeled with the same grayvalue as the wanted objects but which do not belong it. This points are included in the triangulation process and lead to triangles which are hard to eliminate by spatial criteria. Therefore the preprocessing has to take care and extract only pixels of one kind of structured object.

In the three approaches, the extraction of regions characterized by the repetition of symbols does not require the symbols to be complete for including them into the regions, as would be necessary using template matching, for instance. All methods were not able to separate regions covered with different figures, since they only analyse the distance between pixels and not the meaning of each figure or symbol.

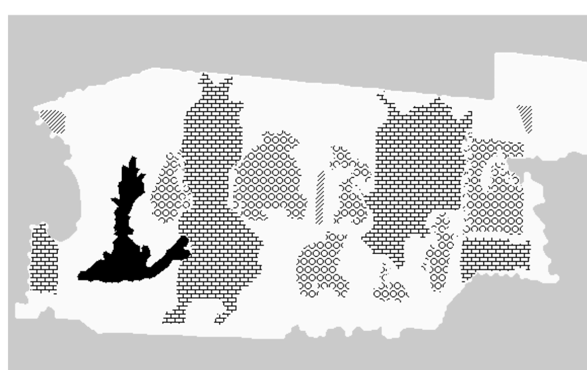

Figure 11: Segmented image

Figure 11 shows the segmented regions using the triangulation approach. Some errors produced in the color segmentation step could be eliminated, but the presence of such errors near the borders of the regions have great influence on the quality of the contours. Improvements could be achieved by improving the quality of the colour segmentation, which would reduce such errors. Nevertheless, the problem of having other figures or symbols close to the textured regions remains a potential source of disturbances.

7 CONCLUSIONS

The inclusion of spatial information into image analysis to extract spreaded elements in a digital image was discussed. Three approaches were compared, the first based on the co-occurrence matrix, another based on the watershed algorithm and the last one using a Delaunay triangulation. The results are different, because of the different principles these methods are based on.

A comparison of the contours of the regions that group spreaded pixels into larger regions reveals that the triangulation based method produces the best results. The computation of the triangulation is more expensive, this could be a disadvantage when dealing with large images. Nevertheless, the estimated contours describe best the regions formed by spreaded pixels or figures. The quality of the colour segmentation plays a decisive role in the quality of the results, since it is responsible for the occurrence of noise or the elimination of valid information prior to the spatial analysis.

References

Doermann, D., 1998. An introduction to vectorization and segemntation. In Tombre, K. and Chabra, A., editors, *Graphic recognition: algorithms ans systems, volume 1389 of Lecture Notes in computer science*, pages 1-7, Berlin, Springer

Haralick, R. and Shapiro, L., 1997. Computer and Robot Vision, volume 1. Addison-Wesley Publishing Company.

Hawkins, J., 1970. Textural properties for pattern recognition. In Lipkin, B. and Rosenfeld, A., editors, Picture processing and psychopictorics, Lecture Notes in computer science, pages 347-370, New York, Academic Press.

Jenson, S. and Domingue, J., 1988. Extraction of topographic structure from digital elevation data for geographic information system analysis. Photogrammetric Engeneering and Remote Sensing, 54(11):1593-1600.

Kunz, D., Schilling, K., Vögtle, T., 1997. A new approach for sattelite image analysis by means of a semantic network. In Förstner, W. and Plümer, L., editors, SMATI 97, pages 20-36, Birkhäuser, Basel.

Mark, D., 1983. Automated detection of drainage networks for digital elevation models. In Proceedings of Auto-Carto 6, volume 2,Ottawa Ontario, Canada

Niemann, H., 1974. Methoden der Mustererkennung, vol 2, Akademische Verlagsgesellschaft, Frankfurt am Main.

Okabe, A., Boots, B. and Sugihara, K., 1992. Spatial Tessellations: Concepts and Application of Voronoi Diagrams. John Willey Press, Chichester.

Pratt, W., 1974. Digital Image processing. John Willey and Sons, New York.

Silva Centeno, J.,Kishi, R.T. and Bähr H.-P., 1999, Recognition of buildings using scanned maps and laser scanner altitude data. PFG, Journal of the DGPF, Vol 1/99, E, Schweizbart'sche Velagsbuchhandlung, Stuttgart.

Vincent, L. and Soille, P., 1991. Watershed in digital spaces: An efficient algorithm on immersion simulation. IEEE-PAMI, 13(6):583-589.

Weindorf, M., 1994. Segmentation von Siedlungsflächen aus Satellitenbilddaten. Diploma thesis, Karlsruhe University.