Generalization of Linear Patterns Based on MDL Criterion

He-Ping Pan and Wolfgang Förstner
Institut für Photogrammetrie, Universität Bonn
Nussallee 15, 53115 Bonn 1, Germany
Email: heping@ipb.uni-bonn.de

Abstract

A domain-independent objective mechanism is developed for generalization of linear patterns. It is based on the Minimum-Description-Length principle, seeking the simplest description of a given polyline. The hypotheses are generated by the farthest point algorithm. The whole mechanism is objective in the sense of without using any control parameter. This mechanism has been tested on segmented images and polygon maps.

1 Introduction

Generalization of linear patterns is a common problem in computer vision, computer graphics, and cartography. The generalization is always guided and constrained by certain criteria for certain application purposes. For example, in polyhedral scenes, long straight lines are preferred which should be produced or retained after the generalization. One of the popular and effective algorithms is the recursive splitting according to maximum distance developed simulataneously by [2] in computer vision and by [1] in geoinformatics. This algorithm makes use of a control parameter: a threshold for the minimum distance to stop the recursive splitting process. This threshold is subjective and dependent on application situations. For a given situation, the threshold is then fixed for all the splittings, which yields very different effects to our perception.

Our approach is to use the Minimum-Description-Length (MDL) criterion to seek the simplest description of a given linear pattern, which produces the most probable significant polyline for a given linear pattern.

2 Problem Formulation of Linear Pattern Generalization

Let [ ] denote a series (an ordered point set) which correponds to a list, { } denote a set, and $\{x\}$ denote the set consisting of the elements of a series x. A linear pattern L is represented by a series of points $L = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \dots \quad \mathbf{p}_n]$. Each point $\mathbf{p}_i$ is defined by a coordinate pair $(x_i, \quad y_i)$. As we assume L is extracted from an image without any generalization, so every pair of neighboring points $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$ in the series are actually connected in the space (image), but two points $\mathbf{p}_i$ and $\mathbf{p}_j$, $\mid j-i \mid > 1$ that are not neighboring in the series are generally also not neighboring in the space. Thus, the whole series forms a linear pattern which is a polyline consisting of limited straight line segments $[\mathbf{p}_1 \quad \mathbf{p}_2]$, $[\mathbf{p}_2 \quad \mathbf{p}_3]$, ..., $[\mathbf{p}_{n-1} \quad \mathbf{p}_n]$.

The goal of generalization is to produce a new series $G = [\mathbf{p}'_1 \quad \mathbf{p}'_2 \quad \dots \quad \mathbf{p}'_m]$, where $m \leq n$, and each $\mathbf{p}'_i$ is a function of the original point set $\{L\}$. Intuitively, G with a minimal amount of data should preserve the perceptual significance of L with the minimal loss of information contained in L. Previously, actual criteria were dependent on subjective satisfaction for given applications.

The basic reason for generalization of a polyline L into a simpler one G is that we assume there are positional noise N for the internal points of L. Here we use symbol N to denote the series of those internal points that cause lines to look noisy:

$$\{L\} = \{G\} + \{N\} \tag{1}$$

Here we use $+$ and $-$ to denote the union of two sets, and the substraction of a set from another set.

Our basic strategy to generate a hypothesis G is to select significant points from L, which will then form G. The other points of L that are not selected are treated as the contents of the noise N. Therefore, G and N can be expressed as follows:

$$\begin{aligned} G &= [\mathbf{p}_{i_1} \quad \mathbf{p}_{i_2} \quad \dots \mathbf{p}_{i_m}] \\ &\quad \mathbf{p}_{i_k} \in L, \quad i_k \in \{1,2,...,n\}, \\ &\quad k \in \{1,2,...,m\}, \quad i_1 = 1, \quad i_m = n \end{aligned} \tag{2}$$

$$\{N\} = \{L\} - \{G\} \tag{3}$$

$$\begin{aligned} N = [&\mathbf{p}_{i_1+1} \quad \mathbf{p}_{i_1+2} \quad \cdots \quad \mathbf{p}_{i_2-1} \\ &\mathbf{p}_{i_2+1} \quad \mathbf{p}_{i_2+2} \quad \cdots \quad \mathbf{p}_{i_3-1} \\ &\cdots \\ &\mathbf{p}_{i_{m-1}+1} \quad \mathbf{p}_{i_{m-1}+2} \quad \cdots \quad \mathbf{p}_{i_m-1} \quad] \end{aligned} \tag{4}$$

where k is the index for a point $\mathbf{p}_{i_k}$ in G, while this point is indexed with i_k in L. $i_1 = 1$ and $i_m = n$ means the starting and ending point should be retained, because each such point has a topological role in the polyline network.

The original recursive splitting algorithm of Duda and Hart and all later variations of others use subjective control parameters for the transform from L to G. Our approach is to maximize the aposterior probability of the data set L taking the generalized pattern G as a model. A more general form of this Bayesian criterion is the Minimum-Description-Length (MDL)

criterion [7] which enables us to combine variables of different models in a uniform measure — the number of bits. Let us suppose there be a set of generalized patterns $\mathcal{G} = \{G_k \mid k = 1, 2, ..., \}$, each G_k is an alternative generalization of the original pattern L. Let $l(x)$ denote the description length of a variable x. The joint description length of L and G as a model of L is $l(L, G)$,

$$l(L, G) = l(L \mid G) + l(G) = l(N) + l(G) \tag{5}$$

The MDL criterion selects the best generalization G with two conditions:

The necessary condition:

$$l(L, G_k) \quad < \quad l(L), \qquad G_k \in \mathcal{G} \tag{6}$$

where $l(L)$ is the description length of L without any model. The sufficient condition:

$$l(L, G) \quad = \quad \min\{l(L, G_k) \mid G_k \in \mathcal{G}\} \tag{7}$$

With this criterion, the goal of generalization is to seek a polyline G that corresponds to the minimization of the joint description length $l(L, G)$.

3 Encoding Polyline and Noise

We now address the problem of how a polyline L and a generalization G as well as the noise N can be encoded in an optimal way. In this section, we describe two random Markov chain models, one for polylines L and G, and another for the noise N.

3.1 A Random Markov Chain Model for Vector Coding

A direct and efficient encoding for a polyline L is vector coding: A starting point $\mathbf{p}_1 = (x_1, y_1)$ is given. With index i starting from 2, each next point is encoded with a position difference vector $\mathbf{d}_i = (u_i, v_i)$ stemming from its previous point $\mathbf{p}_{i-1} = (x_{i-1}, y_{i-1})$, $\mathbf{p}_i = \mathbf{p}_{i-1} + \mathbf{d}_i$.

The length l of this encoding for the polyline L

$$l(L) = l(\mathbf{p}_0) + \sum_{i=1}^{n} l(\mathbf{p}_i) = l(\mathbf{p}_0) + \sum_{i=1}^{n} l(\mathbf{d}_i) \tag{8}$$

Let us use the isomorphism between a complex number and a vector in 2-dimensional space,

$$\mathbf{d}_i = \mathbf{p}_i - \mathbf{p}_{i-1} = u_i + jv_i = s_i \cdot e^{j\phi_i} \tag{9}$$

where $j = \sqrt{-1}$ and

$$s_i = \sqrt{u_i^2 + v_i^2}, \quad \phi_i = \arctan\frac{v_i}{u_i} \tag{10}$$

According to the MDL principle, the description length of this vector is

$$l(\mathbf{d}_i) \quad = \quad -lb\, p(\mathbf{d}_i) \tag{11}$$

where lb is the logarithm with base 2, $p(\mathbf{d}_i)$ is the probability of $\mathbf{d}_i$ which is to be computed according to a stochastic model described as follows.

Because each $\mathbf{d}_i$ is a position difference vector of a point $\mathbf{p}_i$ relative to its previous point $\mathbf{p}_{i-1}$, therefore in order to encode this kind of vectors in an efficient way, we must have a stochastic model which describes the variation of the length s_i over s_{i-1}, and the angle ϕ_i over ϕ_{i-1}, with $i = 3, \quad 4, \quad ..., \quad n$.

Consider the ratio between $\mathbf{d}_i$ and $\mathbf{d}_{i-1}$,

$$\frac{\mathbf{d}_i}{\mathbf{d}_{i-1}} = \frac{s_i e^{j\phi_i}}{s_{i-1} e^{j\phi_{i-1}}} = e^{r_i + j\alpha_i} \tag{12}$$

where

$$r_i = \ln\frac{s_i}{s_{i-1}}, \quad \alpha_i = \phi_i - \phi_{i-1} \tag{13}$$

To encode $\mathbf{d}_i$, as $\mathbf{d}_{i-1}$ is already encoded, so we only need to encode the differential part, The description length of $\mathbf{d}_i$ can then be formulated as

$$l(\mathbf{d}_i) = -lb\, p(\mathbf{d}_i) = -lb\, p(r_i, \alpha_i) = -lb\, p(\mathbf{g}_i) \tag{14}$$

where $p(r_i, \alpha_i)$ is the joint probability of r_i and α_i, and $\mathbf{g}_i$ denotes $(r_i \quad \alpha_i)^T$.

With this expression, it is now clear that $l(\mathbf{d}_i)$ is a markov chain with the assumption that the logarithm of the length ratio r_i and the difference of the directions α_i are stochastic. In general, we assume they follow the normal distribution

$$\mathbf{g}_i = (r_i \quad \alpha_i)^T \sim \mathbf{N}(\mathbf{0}, \mathbf{C}) \tag{15}$$

where $\mathbf{C}$ is the covariance matrix. The probability density function of $\mathbf{g}_i$ is

$$p(\mathbf{g}_i) = \frac{1}{2\pi\sqrt{|\,\mathbf{C}\,|}} \exp\left(-\frac{1}{2}\mathbf{g}_i^T\, \mathbf{C}^{-1}\, \mathbf{g}_i\right) \tag{16}$$

We further assume the two variables are independent, so the covariance matrix $\mathbf{C}$ is diagonal, which may be estimated from the given data. Conversely, if these parameters are given, a class of polylines can be generated from this random Markov chain model.

Let $\hat{x}$ denote the estimated value of x. According to formula (14), the description length $l(\mathbf{d}_i)$ is computed as

$$l(\mathbf{d}_i) = \frac{1}{\ln 2}\left(\ln(2\pi\hat{\sigma}_r\hat{\sigma}_\alpha) + \frac{1}{2}((\frac{r_i}{\hat{\sigma}_r})^2 + \left(\frac{\alpha_i}{\hat{\sigma}_\alpha}\right)^2)\right) \tag{17}$$

According to formula (8), the total description length of the original polyline L is

$$l(L) = l(\hat{\sigma}_r) + l(\hat{\sigma}_\alpha) + l(\mathbf{p}_1) + l(\mathbf{d}_2) + \sum_{i=3}^{n} l(\mathbf{d}_i) \tag{18}$$

$$\sum_{i=3}^{n} l(\mathbf{d}_i) = \frac{n-2}{\ln 2}(\ln(2\pi) + \ln\hat{\sigma}_r + \ln\hat{\sigma}_\alpha + 1) \tag{19}$$

For details of the computation of each term in this formula, see [6]. G is also a polyline, the description length of G is also to be computed with the formula (18) with G replacing L.

3.2 A Random Markov Chain Model for Positional Noise

Linear patterns including lines and edges extracted from images are comtaminated by image noise, round-off errors of rasterization (non-optimal resolution), and imperfect low-level algorithm for edge/line detection or region segmentation. Because we only consider the polylines, we assume a polyline G as an ideal linear pattern is transformed to a detected rasterized linear pattern L which contains positional noise for each internal pixel. As G consists of a number of line segments $[\mathbf{p}_{i_1} \quad \mathbf{p}_{i_2}]$, $[\mathbf{p}_{i_2} \quad \mathbf{p}_{i_3}]$, ..., $[\mathbf{p}_{i_{m-1}} \quad \mathbf{p}_{i_m}]$, so the generalization G and the noise N can be decomposed as

$$\begin{aligned} G &= \diamond_{k=1}^{m-1} G_k \qquad (20) \\ & \quad G_k = [\mathbf{p}_{i_k} \quad \mathbf{p}_{i_{k+1}}] \\ N &= \diamond_{k=1}^{m-1} N_k \qquad (21) \\ & \quad N_k = [\mathbf{p}_{i_k+1} \quad \mathbf{p}_{i_k+2} \quad \cdots \quad \mathbf{p}_{i_{k+1}-1}] \end{aligned}$$

where we use $\diamond$ to denote the concatenation of two or more series. With this decomposition, N_k can be considered as the noise to G_k. Because G_k is supposed to be an ideal line segment and N_k is a rasterized series, so the positional noise N_k takes the form of shift of each internal point in the direction perpendicular to G_k, because the shift along G_k is negligible to the shape of G_k.

In order to consider these shifts, we need to first transform the coordinates of points of N_k into a local reference system taking G_k as the x-axis direction. Then, ideally the y-coordinate for each internal point of N_k should be zero if there is no noise. The positional noise takes the form of non-zero y-coordinates. Generally the positional noise for two neighboring points are correlated. It is obvious that the difference between two neighboring y_{t-1} and y_t is generally smaller than $\mid y_{t-1} \mid$ and $\mid y_t \mid$, so it is cheaper to only encode such differences. This kind of encoding may lead to a Random Markov Chain model for the positional noise along L_i:

$$y_{j+1} = a\, y_j + \varepsilon_k, \quad (j = i_k, i_k+1, ..., i_{k+1}-1) \qquad (22)$$

where a is a parameter of correlation, ε_k is a variable (error term) which follows a normal distribution:

$$\varepsilon \sim N(0, \sigma_k) \qquad (23)$$

where σ_k is to be estimated from the given data N_k,

$$\hat{\sigma}_k = \sqrt{\frac{\sum_{j=i_k}^{i_{k+1}-1} \varepsilon_j^2}{i_{k+1} - i_k}} \qquad (24)$$

where $\varepsilon_j = y_{j+1} - a\, y_j$.

With $a = 0$ the noise is uncorrelated, which refers to roughness. It is equal to say that the $y_j's$ themselves follow a normal distribution. With $a = 1$ the noise is correlated, which refers to smoothness. In general, estimated $\hat{a}$ is between $(0, 1)$. Because we use a fixed description length of this parameter, so a needs not to be encoded.

Therefore, the description length for each internal point $\mathbf{p}_j$ is

$$\begin{aligned} l(\mathbf{p}_j) &= -lb\, p(\mathbf{p}_j) = -lb\, p(y_j) \\ &= -lb \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_k} \exp\left(-\frac{\varepsilon_j^2}{2\hat{\sigma}_k^2} \right) \right) \\ &= \frac{1}{\ln 2} \left(\frac{1}{2} \ln(2\pi) + \ln \hat{\sigma}_k + \frac{\varepsilon_j^2}{2\hat{\sigma}_k^2} \right) \qquad (25) \end{aligned}$$

The description length of N_k is then

$$\begin{aligned} l(N_k) &= l(\hat{\sigma}_k) + \sum_{j=i_k}^{i_{k+1}} l(\mathbf{p}_j) = l(\hat{\sigma}_k) + l(y_{i_k}) \\ &\quad + \frac{i_{k+1} - i_k}{2 \ln 2} (\ln(2\pi) + 2 \ln \hat{\sigma}_k + 1) \qquad (26) \end{aligned}$$

For $l(y_{i_k})$, we should assume it is a random in the whole range of y's in N_k: $[y_{min}^{(k)} \quad , y_{max}^{(k)}]$, $l(y_{i_k}) = -lb\, p(y_{i_k}) = lb\, (y_{max}^{(k)} - y_{min}^{(k)})$. Let $\hat{\sigma}_0$ be the resolution of $\hat{\sigma}_k$ is, then $l(\hat{\sigma}_k) = lb\, (\frac{\hat{\sigma}_k}{\hat{\sigma}_0})$. The total description length N given G as a hypothesized generalization of L is then

$$l(N) = l(L \mid G) = \sum_{k=1}^{m-1} l(N_k) \qquad (27)$$

4 A Recursive Mechanism for Hypothesis Generation and Minimization of Objective Function

Our objective can now be clearly defined: We are seeking a generalization G such that

$$l(L, G) = l(L \mid G) + l(G) = l(N) + l(G) \longrightarrow \min \qquad (28)$$

This is our objective function to minimize. The problem is two-fold: at one hand, we need to generate alternative hypotheses of generalization; at the other hand, we need to select better hypotheses with shorter joint description length $l(L, G)$ in order to reach the minimum of $l(L, G)$. To solve this two-fold problem, we will use the recursive spliting mechanism similar to that proposed by Duda and Hart, but with a different but well principled criterion. In the following, we will describe three significant hypotheses and associated joint description lengths.

For generality, let us suppose we are considering a portion $L_{i,j}$ of an original linear pattern L: $L_{i,j} = [\mathbf{p}_i \quad \mathbf{p}_{i+1} \quad \cdots \quad \mathbf{p}_j]$, where $j - i \geq 2$. When $i = 1$, and $j = n$, the $L_{i,j}$ refers to the original series L.

Hypothesis 0:

It is obvious that the simplest generalization $G_{i,j}^{(0)}$ of $L_{i,j}$ is

$$G_{i,j}^{(0)} = [\mathbf{p}_i \quad \mathbf{p}_j] \qquad (29)$$

without any internal points. In this case, the total description length of $L_{i,j}$ with $G_{i,j}^{(0)}$ as a model is

$$l_0(L_{i,j}) = l(G_{i,j}^{(0)}) + l(N_{i,j}^0) \quad (30)$$

$$l(G_{i,j}^{(0)}) = l(\mathbf{p}_i) + l(\mathbf{p}_j) \quad (31)$$

where $N_{i,j}^{(k)}$ denote the positional noise in $L_{i,j}$, given $G_{i,j}^{(k)}$ as a model of $L_{i,j}$.

Alternative Hypothesis 1:
Alternative hypotheses for generalization can be generated in many different ways. However, the most objective way is to split this series at an internal point $\mathbf{p}_s$, $i+1 \leq s \leq j-1$ into two subseries $L_{i,s}$ and $L_{s,j}$, where index s is determined by

$$\mid y'_s \mid = \max\{y'_k \mid k = i+1, ..., j-1\} \quad (32)$$

i.e. $\mathbf{p}_s$ is the farthest point from the line defined by $\mathbf{p}_i$ and $\mathbf{p}_j$. Notice $\mathbf{p}_s$ can be in either side of the line $G_{i,j}^{(0)}$. In this case, the generalized series at this stage is

$$G_{i,j}^{(1)} = [\mathbf{p}_i \quad \mathbf{p}_s \quad \mathbf{p}_j] \quad (33)$$

The description length l_1 of $L_{i,j}$ by taking $G_{i,j}^{(1)}$ as a model is

$$l_1(L_{i,j}) = l(N_{i,j}^{(1)}) + l(G_{i,j}^{(1)}) \quad (34)$$

$$l(G_{i,j}^{(1)}) = l(\mathbf{p}_i) + l(\mathbf{p}'_s) + l(\mathbf{p}_j) \quad (35)$$

where $\mathbf{p}'_s$ is the new coordinate of $\mathbf{p}_s$ in the local coordinate system taking $[\mathbf{p}_i \quad \mathbf{p}_j]$ as the x-axis.

Alternative Hypothesis 2:
If we consider the points of $N(0)_{i,j}$, they may appear at both sides of the line $G^{(0)}$, there could be a maximum value and a minimum value of the new y-coordinates at two different sides. Let $\mathbf{p}_{s_1}$ and $\mathbf{p}_{s_2}$ with $s_1 < s_2$ denote the two extremum points (we do not disinguish which one is the maximum or minimum) with the condition:

$$y_{s_1} \cdot y_{s_2} < 0 \quad (36)$$

So, the second alternative hypothesis is

$$G_{i,j}^{(2)} = [\mathbf{p}_i \quad \mathbf{p}_{s_1} \quad \mathbf{p}_{s_2} \quad \mathbf{p}_j] \quad (37)$$

If the condition (36) is not satisfied, then $G_{i,j}^{(2)}$ does not exist.

The description length l_1 of $L_{i,j}$ by taking $G_{i,j}^{(2)}$ as a model is

$$l_2(L_{i,j}) = l(N_{i,j}^{(2)}) + l(G_{i,j}^{(2)}) \quad (38)$$

$$l(G_{i,j}^{(2)}) = l(\mathbf{p}_i) + l(\mathbf{p}'_{s_1}) + l(\mathbf{p}'_{s_2}) + l(\mathbf{p}_j) \quad (39)$$

where $\mathbf{p}'_{s_1}$ and $\mathbf{p}'_{s_2}$ are the new coordinate of $\mathbf{p}_{s_1}$ and $\mathbf{p}_{s_2}$ in the local coordinate system taking $[\mathbf{p}_i \quad \mathbf{p}_j]$ as the x-axis.

Dropping the equal terms in (30), (34), and (38), these formulas can be rewritten as

$$l_0(L_{i,j}) = l(N_{i,j}) \quad (40)$$

$$l_1(L_{i,j}) = l(\mathbf{p}'_s) + l(N_{i,j}^{(1)}) \quad (41)$$

$$l_2(L_{i,j}) = l(\mathbf{p}'_{s_1}) + l(\mathbf{p}'_{s_2}) + l(N_{i,j}^{(2)}) \quad (42)$$

where each $\mathbf{p}'_k$, $k = s, s_1, s_2$ is supposed to be a random in the bounding box $(w_{i,j}, h_{i,j})$ (for width and height) of all the points between $\mathbf{p}_i$ and $\mathbf{p}_j$,

$$l(\mathbf{p}_k) = lb\, w_{i,j} + lb\, h_{i,j} \quad (43)$$

The decision is now to select one from the three hypotheses with smallest description length l. If the hypothesis 0 is selected, then stop for this local $L_{i,j}$, otherwise, $L_{i,j}$ is splitted into two or three new subseries.

Starting from $i = 1$ and $j = n$, the original series L is first splitted into two or three sub-series. Each new subseries can be tested again for further splitting. The test with a selection of the hypothesis 0 is a hard criterion to stop a local split. This recursive splitting of $L_{i,j}$ is a gradient decendent approach to reach the minimization of the total description length of the original series L together with the final series G as the most probable model. The final generalized series G is an ordered collection of the original starting point $\mathbf{p}_1$, all the accepted splitting points, and the original ending point $\mathbf{p}_n$. Finally, the necessary condition (6) can be tested. If it is true, then finally accept the final generalization G. A straightforward implementation of the mechanism described above is a recursive algorithm with the computational complexity $O(n \log n)$. It is also possible to design a nonrecursive algorithm for implementing the recursive mechanism for generalization, but the complexity should be the same as the recursive one.

5 Experiments and Further Remarks

We have tested this mechanism on linear edges extracted from raster images (see Fig. (1)-(3)). In order to study the influence of image noise, we have generated a polygon map by using a generic structural model [5], thus providing an ideal edge map. Then, a raster image is generated from this maps by randomly assigning a gray value to a polygon, and adding a signal-dependent noise to each pixel. The simulated image is segmented through a general MDL-based region-merging algorithm. The crack edges are then extracted. Notice that all the crack edges form a complete network, so each polygon is completely surrounded by a series of crack edges. In case of uncorrelated noise model with $a = 0$, in Fig.2, we can see more knoting points are retained from L into G. In case of correlated noise model with $a = 1$, in Fig.3, the extracted edges are more generalized, with less details.

Because we use polyline-split mechanism for hypothesis generation, a generalized polyline is defined by an ordered retained points of the original linear pattern. Therefore, compared with the ideal edge map, the generalized polylines do not exactly coincide on the original vectorized edges. A postprocessing would be to fit each generalized polyline to its real edge position by least-square method [4]. This work is beyond the scope of pure generalization of provided linear patterns.

6 Conclusions

The mechanism described here for generalization of linear patterns is objective because it is based on MDL criterion, thus without using any control parameter. It is also efficient, because it is implemented in a recursive algorithm, with complexity of $O(n \log n)$. This mechanism and criterion used can be further applied with other approaches which can generate hypotheses of generalization in different ways. However, before applying this mechanism to any specific problem domain, the validity of the models proposed here for linear patterns needs to be verified. Once the proper valid models are available, this mechanism can be easily adapted.

Acknowledgement

This work has been supported by the Deutsche Forschungsgemeinschaft.

References

[1] D. H. Douglas and T. K. Peuker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *The Canadian Cartographer* 10(2):112-122, 1973.

[2] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* John Wiley and Sons, New York, 1973.

[3] W. Förstner. Image analysis techniques for digital photogrammetry. *Schriftenreihe Heft 13, Institut für Phtogrammetrie Stuttgart*, 1989.

[4] C. Fuchs, T. Lcherbach, H.-P. Pan, and W. Frsrner. Land use mapping from remotely sensed images. *Proc. 2nd International Colloquium on Urban Spatial Information Systems*, Wuhan Technical University of Surveying and Mapping, Wuhan, China, 1993.

[5] H.-P. Pan and W. Förstner. Polygon Map Grammars: A generic model for understanding landuse maps and images in remote sensing. W. Förstner and S. Ruwiedel (Eds.): *Robust Computer Vision*, Wichmann Verlag, 1992.

[6] H.-P. Pan and W. Förstner. Generalization of linear patterns based on MDL criterion. Technical report TB-ipb-94/1, Institut fr Photogrammetrie, Universitt Bonn, 1994.

[7] J. Rissanen. Stochastic complexity in statistical inquiry. *Series in Computer Science*, Vol. 15, World Scientific, Singapore, 1989.

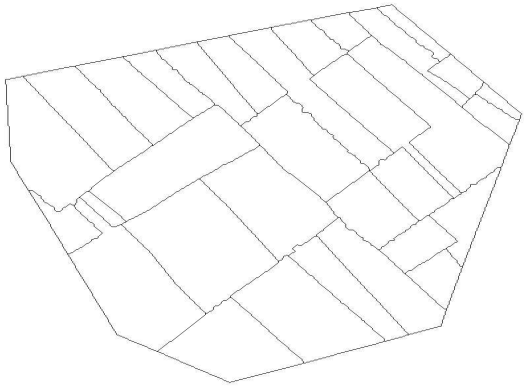

Figure 1: Extracted edges from a simulated polygon map image

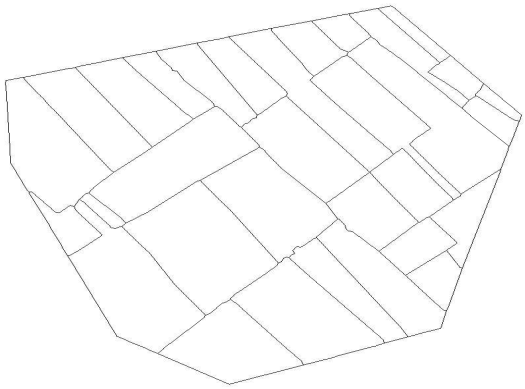

Figure 2: Generalization under uncorrelated noise model

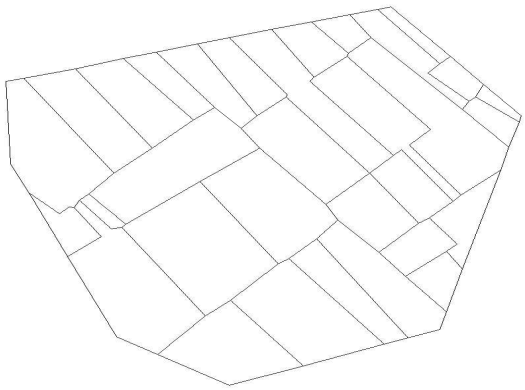

Figure 3: Generalization under correlated noise model