Intelligent Image Databases For Nature Management Optimization

Nelley M. Kovalevskaya[1]

Abstract

Intelligent image databases have to facilitate analyses of space and aerial observations of the environment processes and phenomena for estimation the environment current state. Such image databases should efficiently combine the experts' image interpretations and the corresponding environment knowledge. The paper demonstrates a new application of computer vision to image databases - the use of image texture for annotation, the description of content. The goal was to use a scheme that is able to automatically decide on the image features and based upon psychophysical studies of human perception nature and computer vision models in contrast to multiple cue-based schemes being still heuristic. The approach provides a learning algorithm for selecting the most representative features of the homogeneous and piecewise-homogeneous data. Highly specializes and context-dependent features are extracted automatically and spatial information is preserved.

1. Introduction

Environmental degradation is an inseparable part of our civilization's development. Although people are more aware of what causes the degradation and take preventive steps, the consequences of degradation are becoming more and more visible. Efficient environmental policies are obstructed due to scarcity of complete and confident area information, drawbacks of conventional field data acquisition and insufficient environmental information management in the absence of powerful digital systems.

The good prospects for joint application of space and aerial images for studying and observing environmental processes were found at the very beginning of the space age. Such imagery permits us to cover environmental processes on larger territories and explore their development. Nevertheless, the collected images were not widely used in practice because those taken from space were underestimated by professionals. This was also due in part to the absence of systematic image databases

[1] Institute for Water and Environmental Problems, Papanintsev St., 105; Barnaul Russia 656099; E-mail: knm@iwep.secna.ru, knm@iwepsb.altai.su

that allow us to investigate particular environmental disasters and to learn how to recognize one or another environmental process on land surfaces.

At present, ecologists are deterred by their minimum contribution to the development of means and technologies for acquiring and applying remotely sensed data to environmental problems. Space images are mostly collected without regard for the requirements of this end user group. Since remote sensing is being left in the hands of technocrats, it has frequently no chance of being the powerful practical ecological tool it could be given its abilities to provide for the permanent imaging of the Earth's surface on a day-by-day basis. In spite of these abilities, the potential for environmental degradation assessment by means of remote sensing has not yet been realized.

There is a need to involve environmental experts in space planning and underlying aerial observations to investigate particular environmental processes, as well as in analysis and systematical collection of relevant images. This will facilitate the use of these results in practice for detecting and assessing environmental degradation.

2. Main Issues

Intelligent image databases (IDBs) have to facilitate thorough analyses of space and aerial observations of the environment processes and phenomena for evaluation of the current state of the environment. Such image databases should efficiently combine the experts' image interpretations and the corresponding environmental knowledge. An environmental knowledge-based approach to building an intelligent IDB suggests that the IDB should be created on the basis of a special image model to be able to describe, generate, store and retrieve specific samples of environmental degradation by their visual and semantic contents.

Experiments have shown that it is very difficult to characterize complex environmental objects by their images, so that often it is done through qualitative and descriptive terms, the nomenclature being verbal and non-unique. For example, fuzzy adjectives like "hilly", "rough", and "flat" are used to describe typical landforms, and terms like "fine", "coarse", "close", "loose", "plain" and "twilled" distinguish between different image patterns of the landscape (Gimel'farb and Kovalevskaya 1995; Kovalevskaya 1998). Although the qualitative terms are a useful tool for human descriptive interpretation, they involve inherent subjectivity and fuzziness. This is true for any visible object of natural origin, for example patterns of snow cover pollution. It should be noted that, by and large, various groups of experts come to different and, sometimes, contradictory conclusions based on the same environmental image patterns. Thus, it is very important to distinguish such ambiguous environmental patterns in images.

Conversion of various qualitative attributes and descriptions of particular environmental patterns to measurable, quantitative properties offers a wide and open research field and is highly significant for assessing environmental degradation. In general, it boils down to the problem of spatial (visual) knowledge discovery and modeling.

Actually, while IDBs are rather popular, spatial knowledge modeling still faces many problems and challenges. The multiple cue-based schemes generally are still heuristic and not based upon psychophysical studies of the nature of human perception and computer vision models.

It is noticeable that the most enlightening and insightful research into visual knowledge has tended to consider the ideas of one field with reference to another. One can benefit by taking cues from the inferred representations, dynamics and architecture of the human visual systems. Perception might be defined as the extraction of "meaning" from a (visual) stimulus. The goal of a visual knowledge discovery system is to process such stimuli and extract a more meaningful and concise representation. The system must analyze the relationships among elements of an image in order to infer some higher-level abstracted representation.

The problems of visual knowledge modeling in IDB fall into three broad areas: retrieval, segmentation and recognition/classification. Much of the past research in these fields has concentrated on the cue extraction stage. The success of the problem solution is heavily dependent upon the choice of cues upon which human perception is based. Most natural surfaces exhibit texture that appears to be a vital cue for visual knowledge discovery systems. Scenes containing pictures of wood, grass, etc. can be easily classified based on texture rather than shape or color. But a concise definition of texture has eluded past research.

In 1970, an early researcher (Muerle 1970) noted that a "definition of texture does not exist". In the same year, Hawkins (1970) expressed similar sentiments in observing how difficult it is to define "the very concept of texture", especially given the obvious "psychological fact that human beings perceive and recognize textures". Since then numerous different texture definitions have been attempted in the literature, demonstrating the difficulty in defining texture is to measure it (Gimel'farb 1996).

As computing technology appeared to be widely available, researchers began to apply a wide variety of numerical algorithms to describe textural images. In general these algorithms relied upon computing a set of features which could be used to classify a texture by measuring the similarity of its features to that of another. And again the major difficulty lies in the choice of features used to describe textures. If some knowledge is available about the input patterns then the features can be optimally selected to maximize the likelihood of correct matching. Thus, the most important issues of visual knowledge discovery and modeling are the following:

- A model should be based on the study of a common standard, this being the nature of human perception.
- A model should be flexible and be able to extract a concise representation of the immense amount of information.
- A model should be able to automatically decide on the features for knowledge discovery purposes.

From this point of view generative representations appear to be very attractive due to their ability to generate images by a special technique and check the expressibility of the representation directly by comparing the generated images with the learning ones visually and quantitatively.

Julesz & Bergen (1983) review research over the previous 20 years, acknowledging that it has taken this length of time to discover that texture perception depends on local second-order features alone and that global higher-order statistical parameters can be ignored. Research with texture pairs having identical second-order statistics has revealed that the texture discrimination system cannot globally process third- and higher-order statistics, and that discrimination is the result of a few local conspicuous features. It appears that only the second-order statistics have perceptual significance, and the relative phase between pairs of pixels cannot be perceived without detailed scrutiny by local attention.

For a human expert, the visual environmental features of an image relate mostly to specific spatially homogeneous or piecewise-homogeneous textures created by "weaving" specific primitive elements, or micro-patterns (Gimel'farb and Kovalevskaya 1998). Experiments in modeling and segmenting image textures confirmed high potential for exploiting new probabilistic models of Gibbs random fields and related stochastic optimization techniques as effective interactive tools for developing the quantitative visual features of the environmental patterns (Kovalevskaya and Gimel'farb 1998; Kovalevskaya 1999). The models are generative ones and take into account only first-order and second-order statistics or multiple pairwise interactions between the gray levels in the pixels. An effective learning scheme is used to automatically decide on the conspicuous texture features, namely to recover a structure and strength of the interactions using maximum likelihood estimates.

3. Models with Multiple Pairwise Interactions

Visual patterns can be specified as *(i)* grayscale or color visual patterns of natural surfaces and *(ii)* simulated patterns, which approach, within certain limits, these natural patterns.

The visual pattern elements are called *textons* in Julesz 1981. In a finite lattice, each texton is supported by a specific spatial combination of the pixels. Gray levels in these pixels represent the texton so that a particular support produces a set of possible textons.

Let two grayscale images represent the same translation-invariant visual pattern if they have the same or, at least, closely similar marginal probability distributions of gray levels in the pixels and of gray level co-occurrences in the pixel pairs that belong to a particular characteristic subset of the families of translation-invariant pixel-pairs.

Then the self-similarity can be quantitatively measured by using one or another statistical goodness-of-fit test for the sample relative frequency distributions of gray levels and of particular gray level co-occurrences in the images. Gibbs random fields as image models directly exploit this scheme (Gimel'farb and Kovalevskaya 1998; Kovalevskaya 1999).

Probabilistic visual pattern representation is based on the assumption that each digital image is a sample of a particular random field and can be matched in experiments with a particular probability. An image representation relates the image signals (e.g. gray levels) to the probability. Also if the representation is generative then various visual image samples can be simulated in such a way that their sample relative frequencies tend to the probabilities specified by the model.

Let any random field X *with* the components X *(i)* taking values x *(i)* from the same finite signal set **Q** be considered as a probabilistic model of patterns. But to be of practical use the model should provide high probabilities of the desired patterns and zeroth or almost zeroth probabilities of other patterns. Usually the desired images to be modeled constitute only a tiny part of all the possible patterns.

The research considered the following three types of grayscale patterns: spatially homogeneous patterns, region maps and piecewise-homogeneous patterns. The patterns and region maps differs only by the physical meaning of the gray levels *x (i)* in the pixels.

In the general case the Markov/Gibbs model is introduced as follows *(Besag, 1974)*. Let $\boldsymbol{X} = (X(i): i \in \boldsymbol{R})$ be a 2-D lattice Markov random field (MRF) with samples $x = (x(i): i \in \boldsymbol{R}; \; q = x(i) \in \boldsymbol{Q})$, where **R** is the finite 2-D lattice $\boldsymbol{R} = ((m,n): m=0,\ldots,M-1; n=0,\ldots,N-1)$ of the size $|\boldsymbol{R}| = M*N$, and $\boldsymbol{Q} = \{0,1,\ldots, q_{max}\}$ is a finite set of the signal values, or gray levels q in lattice sites, or pixels *(m,n)*.

The theorem of Hammersley and Clifford (Geman and Geman 1984) states that the joint probability distribution P $(\boldsymbol{X}=\boldsymbol{x})$ of the MRF over the samples space $\boldsymbol{SS} = (\boldsymbol{x}: x(i) \in \boldsymbol{Q}; i \in \boldsymbol{R})$ of all possible realizations of the field under the positivity condition (that all sample realizations $\boldsymbol{x} \in \boldsymbol{SS}$ have non-zero probability $P(x)>0$) can be represented in the factorized form:

$$P(\boldsymbol{x}) = Z^{-1} \exp\Big(\sum_{\kappa \subset \boldsymbol{R}} V_\kappa (x(i) = q_i : i \in \kappa; q_i \in \boldsymbol{Q}) \Big),$$

where V...(...) denotes potentials, or non-constant functions of the variables $x(i)$ such that their supports κ are cliques, or complete subgraphs in the neighborhood graph. The potentials describe quantitatively the strength of the signal interaction over each clique. The term Z is the normalizing constant:

$$Z = \sum_{\boldsymbol{x} \in \boldsymbol{SS}} \exp\Big(\sum_{\kappa \subset \boldsymbol{R}} \boldsymbol{V}_{\kappa}(x(i) : i \in \kappa)\Big).$$

It should be particularly emphasized that an interaction between the pixels, or equivalently, between gray levels in the pixels, has no physical meaning. It reflects only the fact that some spatial combinations in a particular pattern are more frequent than others: the less uniform the probability distribution of combinations in the pixels that have a particular spatial arrangement, the stronger the interaction between pixels.

Pattern homogeneity is quantitatively defined in this research in terms of conditional probability distributions of spatial combinations. Mostly the distributions are assumed to be translation invariant, that is, in different parts of a homogeneous pattern that can be superimposed by translation, the sample relative frequencies of gray level combinations are expected to be almost the same.

For describing spatially homogeneous and piecewise-homogeneous patterns the interaction structure seems to be more important then the potentials. So traditional automodels cannot reflect in full measure the characteristic features of visual patterns mainly because of the pre-defined interaction structure. But, they are widely used in practice due to a common opinion that they are no practicable alternatives.

Contrary to this opinion, it is an easy matter to generalize the Gibbs models so as to involve an arbitrary structure of multiple translation invariant pairwise pixel interactions. In a generalized model each pixel appears in several characteristic second-order cliques, and the interaction structure is formed by uniting the corresponding second-order clique families K_α.

$$K = \bigcup_{\alpha \in A} K_\alpha$$

Each second-order clique family $K_\alpha \equiv K\mu_\alpha\nu_\alpha = [(i,j){:}\ i,j \in \boldsymbol{R};\ i\text{-}j = (\mu_\alpha, \nu_\alpha)]$ contains all the translation invariant pixel pairs (i,j) that have a fixed interpixel, or intra-clique shift (μ_α, ν_α); if $i = (m,n)$ then $j = (m - \mu_\alpha, n - \nu_\alpha)$.

There are many natural and artificial patterns that can be modeled adequately by proposed Gibbs models with multiple pairwise interactions (Kovalevskaya 1999).

This specific class of spatially homogeneous and piecewise homogeneous image visual patterns is called stochastic pattern (Gimel'farb 1996). A stochastic pattern has pixels and pixel pairs as supported for (primitive) textons and is completely specified by the structure and strengths of the translation invariant pairwise pixel interactions

A model-based interaction map shows relative contributions of each clique family to the total energy and can be displayed, for a visual analysis, in a grayscale or color form. Examples of the interaction maps for the natural pattern are displayed in Figure 1.

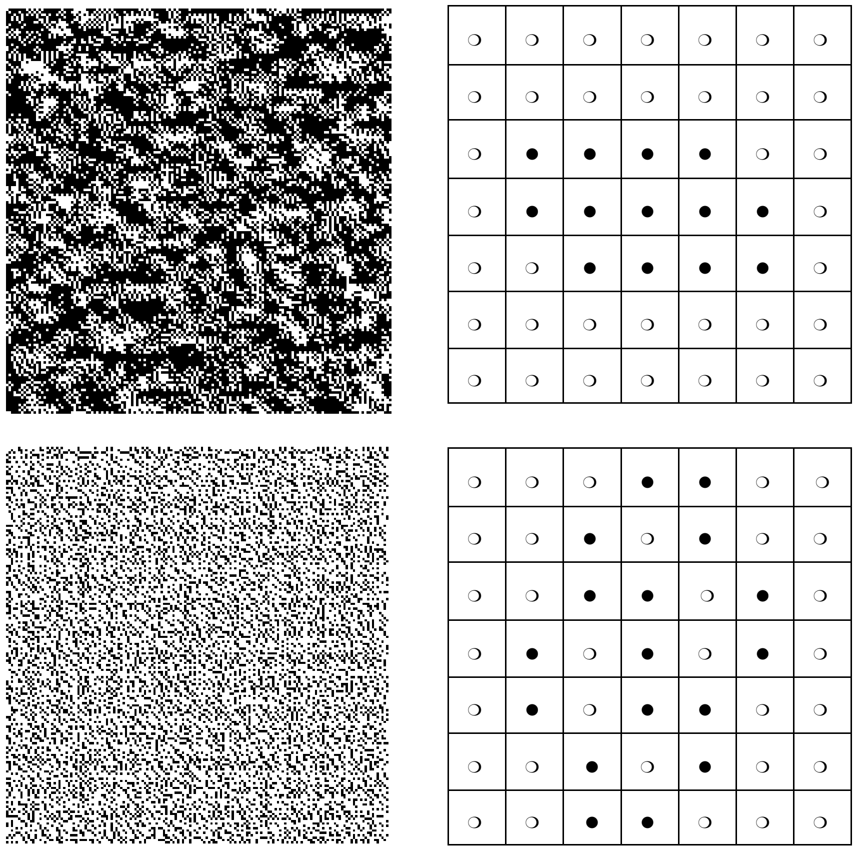
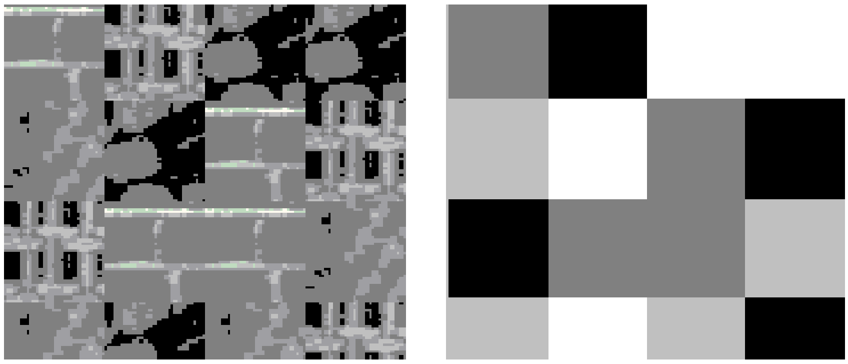
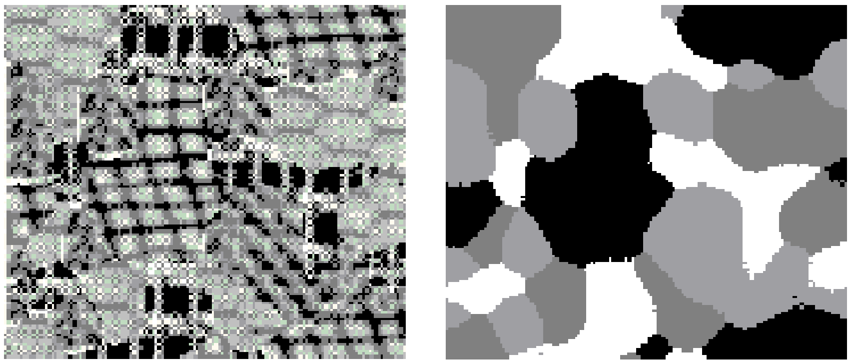


Figure 1: Model-based interaction map to recover most characteristic structure of the pairwise pixel interactions: natural samples and visual forms of interaction maps.

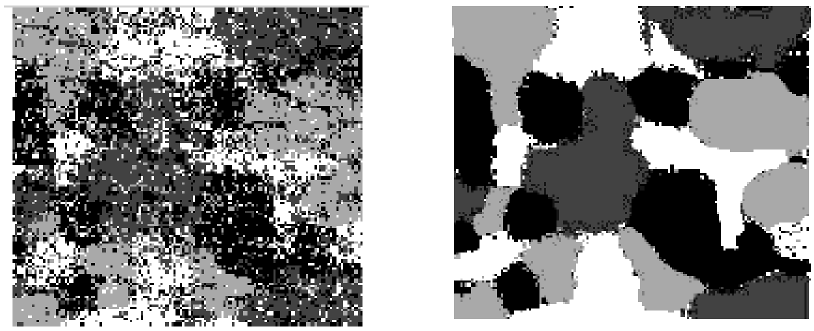
Supervised segmentation is considered as a simulation of a desired region map using a conditional Gibbs model of region maps corresponding to a piecewise-homogeneous grayscale pattern (Figure 2), and the same controllable simulated annealing technique can be used to segment a piecewise-homogeneous image.



(a)



(b)



(d)

Figure 2: Segmentation of the four-region collage of visual patterns: (a) the training collage and its region map, (b) the test collage an ideal region map,(c) segmentation maps (8^{th} and 61^{st} iterations).

Different fragments of the natural image are used to analyze to what extent our model reflects the self-similarity within the patterns (Figure 3). The experiments show that some of the visual patterns really belong to the class of stochastic patterns. In these cases, the natural and simulated patterns possess both a good visual resemblance and high proximity between characteristic families. Although in some cases such proximity does not ensure the visual similarity.

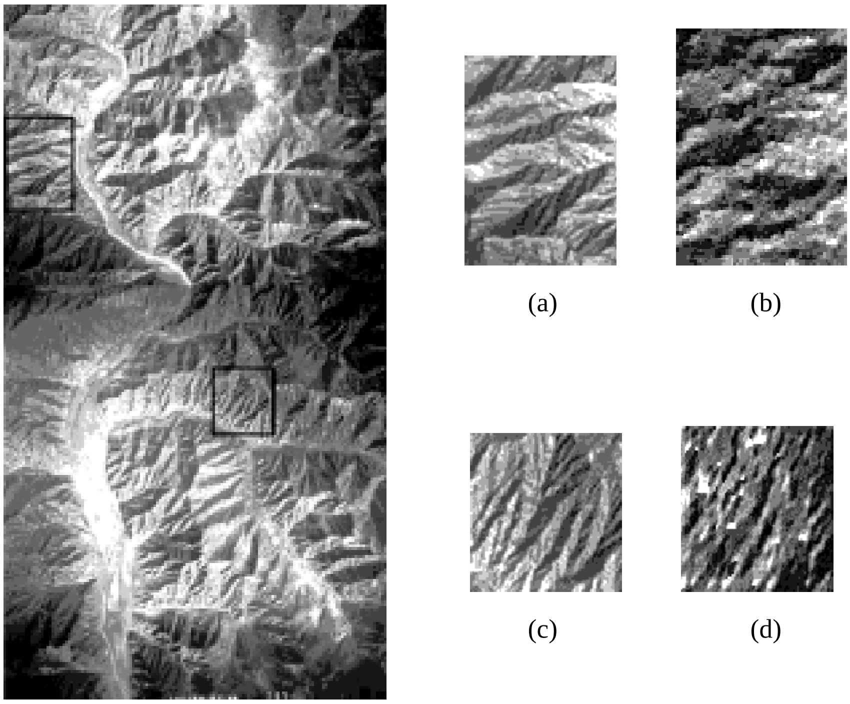


Figure 3: Left image: Terrace valley of Katun' River (Altay mountains, aerial survey).
Right images: Different types of erosion terraces:
(a, c) - training patterns, (b, d) - generated patterns.

These results are applied to investigate structural and functional features of ecosystems. The main attention is focused on:

- the dynamics of land cover, landscape changes, degradation and recultivation processes;
- snow cover pollution due to concentrated industries or single-branch economies in cities;

- infrastructure changes and urbanization;
- National Nature Reserves.

The knowledge-based environmental interpretation of space and aerial images is still more an art than a formal theory. It is mostly descriptive, uses fuzzy terms and is not systematically equated with the measurable attributes. So the next step should be to gain a more penetrating insight into understanding the processes of environmental interpretation, capture the knowledge and, as far as possible, reformulate it in explicit, or quantitative terms.

4. References

Besag, J. E. 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of Royal Statistical Society B36:192-236.

Geman, D., and Geman, S. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6(6): 721 -741.

Gimel'farb, G. and Kovalevskaya, N. 1995. Segmentation of Images for Environmental Studies by Using a Simple Markov/Gibbs Random Field Model. In Proceedings of the Sixth International Conference on Computer Analysis of Images and Patterns, 57 - 64. Prague, Czech Republic: Springer.

Gimel'farb, G. 1996. Texture modeling by multiple pairwise pixel interactions. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(11): 1110-1114.

Gimel'farb, G. and Kovalevskaya, N. 1998. Computational Vision Models for GIS analysis: Intelligent Image Simulation as a Tool for Thematic Mapping, Advances in Geographic Information Systems. In Proceedings of the Sixth International ACM Symposium, 168-169. Washington D.C.: 1999 ACM, Inc.

Julesz, B. 1981. Textons, the elements of texture perception, and their interactions. Nature 290: 91-97

Hawkins, J.K. 1970. Textural properties for pattern recognition. In Lipkin B.S. and Rosenfeld A., editors, Picture Processing and Psychopictorics: Academic Press.

Julesz, B. and Bergen, J.R. 1983. Textons, the fundamental elements of pre-attentive vision and perception of textures. The Bell system Technical Journal 62 (6): 1619-1645.

Kovalevskaya, N. and Gimel'farb, G. 1998. Computational Image Models for GIS Analysis: Texture Modeling as a Tool for Thematic Mapping. In Proceedings of the Forth International Conference on GIS for Nature Management Optimization for Sustainable Development of Territories, 179-190. Barnaul, Russia: Altay State University Publisher.

Kovalevskaya, N. 1999. Gibbs Model of Image As a Tools for Thematic Analysis. Pattern Recognition and Image Analysis 9(2):282-285.

Muerle, J.L. 1970. Some thoughts on texture discrimination by computer. In Lipkin B.S. and Rosenfeld A., editors, Picture Processing and Psychopictorics: Academic Press.