Automatic Interpretation of Multispectral Aerial Photographs

Prof. Dr.-Ing., Dr.-Ing. E.h. Karl-Rudolf Koch
Institut für Theoretische Geodäsie
Rheinische Friedrich-Wilhelms-Universität Bonn
Nußallee 17, 53115 Bonn
Germany

Abstract

The interpretation of digital images is formulated by an analysis at two levels. At the low level the image is segmented based on textures. To the regions thus obtained a meaning is attributed at the high level of the analysis. The segmentation as well as the interpretation are formulated as labeling processes which possess the Markov property so that their distributions follow from Gibbs distributions. The image model and the object model for the interpretation are therefore determined by Gibbs distributions. The results for the interpretation of an aerial photograph of an urban area in the frequency bands red, green and blue are given.

1 Introduction

The interpretation of digital images taken by sensors in aeroplanes or satellites is a task which needs to be solved because of the ever increasing demand for creating or updating maps of different scales of the surface of the earth. The information contained in digital images is not only needed in cartography but also in geodesy and photogrammetry to feed geoinformation systems. The automatic retrieval of the data from the digital images is necessary, as the standard procedure to obtain the information takes too much time. Unfortunately, this part of the field of computer vision is extremely difficult to solve despite of many years of investigations. In the following first results are presented of an automatic interpretation of an aerial photograph taken in the three channels red, green and blue of an urban area with streets, houses, garages, lawns, hedges and bushes.

Interpretation of digital images means to identify objects and attribute a meaning to them, in order to describe the content of the image. Some prior knowledge on the content is needed for the interpretation. This knowledge is collected in the object model which contains the information on the objects and the relations between the objects. The objects, like streets, houses and so on have to be described by their geometry and their functionality so that a semantic model is needed. A relation between objects, which is very helpful for the interpretation, is the neighborhood. It is the context of objects which has to be considered when interpreting digital images.

An image model is also needed, by which the appearance of the objects in the digital image is described. Starting from the pixels, picture primitives have to be extracted to model the objects of the image. This process is generally formulated at different levels. At the low level of the image analysis the pixels are combined to form edges and regions, the picture primitives. At a higher level of the image analysis the primitives may be put together to object primitives which at a high level of the analysis form the objects whose meaning is found by relating them to instances of the object model.

In the following two levels of image analysis are considered based on the assumption that the objects contained in the image differ by their textures. Hence, at the low level a segmentation is applied to gather pixels of identical textures in regions. At the high level of the analysis a meaning is attributed to the regions so that the content of the image can be described.

The segmentation at the low level as well as the interpretation at the high level are defined as labeling processes. At the low level each pixel gets a label by which the region is obtained and at the high level each region gets a label from which the meaning follows. The labeling processes are introduced as random fields which possess the Markov property, that is, the conditional densities of the labels depend on the values of the labels in the neighborhood. By maximizing these densities the values of the labels are estimated. In addition the measurements of the grey values of the three frequency bands are introduced as random variables of a Markov random field as well as the data derived from the regions and used for the interpretation at the high level. As mentioned, prior information on the objects is available so that Bayesian inference is applied.

Markov random fields have been frequently applied for the image processing at the low level, like restauration, edge detection and segmentation, see for instance (KOCH AND SCHMIDT 1994, p.299; PAN 1994). At the high level MODESTINO AND ZHANG (1992) interpreted an image by defining the labels of the objects as a Markov random field. Their work was considerably improved by KÖSTER (1995) who rigorously applied Bayesian statistics and added to the neighbors of objects the indirect neighbors.

Markov random fields have Gibbs distributions according to the theorem of HAMMERSLEY AND CLIFFORD from 1971, see for instance (KOCH AND SCHMIDT 1994, p.261). Gibbs distributions are very flexible to define by means of cliques which are based on neighborhoods so that relations between objects can be easily introduced. But also prior information can be readily incorporated by Gibbs distributions. The main part of the image model for the interpretation applied here is constituted by the texture. It is described, as will be shown, by a Gibbs distribution so that this distribution defines the image model. The object model is also determined by a Gibbs distribution, since the Gibbs distributions introduced for the labeling process at the high level of the image analysis need information from the object model.

In the following chapters the image interpretation is described by considering the analysis at the low level together with the analysis at the high level. The theory for such an aproach has been outlined by KOCH (1995). For easy reference the main results are repeated here together with refinements which were necessary. First results are also presented, they are taken from (KÖSTER 1995).

2 Modified Bayes' Theorem

If θ denotes the vector of unknown parameters, y the vector of observations, the posterior density $p(\theta|y)$ of the parameters θ given the observations y follows from Bayes' theorem by

$$p(\theta|y) \propto p(\theta)p(y|\theta) , \tag{2.1}$$

where $\propto$ denotes proportionality, $p(\theta)$ the prior density and $p(y|\theta)$ the likelihood function, see for instance (KOCH 1990, p.4). Conditional densities will serve as prior densities in the following so that (2.1) needs to be modified. This is readily accomplished by the definition of the conditional density $p(\theta|y,z)$, see for instance (KOCH 1988, p.107),

$$p(\theta|y,z) = \frac{p(\theta,y,z)}{p(y,z)} , \tag{2.2}$$

where z denotes an additional random vector. By applying the definition (2.2) for $p(y|\theta,z)$ we find

$$p(\theta,y,z) = p(\theta,z)p(y|\theta,z)$$

and furthermore

$$p(\theta,z) = p(z)p(\theta|z)$$
$$p(y,z) = p(z)p(y|z) .$$

By substituting these results in (2.2) we obtain a modified Bayes' theorem

$$p(\theta|y,z) \propto p(\theta|z)p(y|\theta,z) , \tag{2.3}$$

since $p(y|z)$ is constant, because y and z are assumed as given. The prior density $p(\theta|z)$ in (2.3) is now definied by a conditional distribution.

The unknown parameters θ are determined by the MAP estimate $\overline{\theta}$ of θ

$$\overline{\theta} = \arg\max_{\theta} p(\theta|y,z) . \tag{2.4}$$

3 Markov Random Fields

As mentioned in the introduction, the image interpreation is solved by a labeling process at the low level and at the high level of the image analysis. At the low level pixels are labeled. Let Ω be the set of pixels

$$\Omega = \{r = (m, n)\ ,\ 0 \le m \le M,\ 0 \le n \le N\}\ , \quad (3.1)$$

where $r \in \Omega$ denotes a pixel at the position (m, n) with M being the maximum number of rows and N of columns of the digital image. The labeling of the pixels is defined as the Markov random field $E(r)$ with value ϵ_r, which determines the texture to which the pixel r belongs,

$$E(r) = \epsilon_r\ ,\ r \in \Omega \quad (3.2)$$

with

$$\epsilon_r = t\ ,\ t \in \mathcal{E}\ ,\ \mathcal{E} = \{1, \ldots, T\}\ , \quad (3.3)$$

where t denotes the label, $\mathcal{E}$ the set of labels and T the number of textures.

Pixels belonging to the same texture form a region representing an object. Thus, at the high level of the image analysis we have the set $\mathcal{K}$ of regions with K elements

$$\mathcal{K} = \{1, \ldots, K\} \quad (3.4)$$

and $p \in \mathcal{K}$ being an element of $\mathcal{K}$. The labeling of the regions is defined as the Markov random field $E(p)$ with value ϵ_p, which determines the meaning of the object represented by the region p,

$$E(p) = \epsilon_p\ ,\ p \in \mathcal{K} \quad (3.5)$$

with

$$\epsilon_p = l\ ,\ l \in \mathcal{E}\ ,\ \mathcal{E} = \{1, \ldots, U\}\ , \quad (3.6)$$

where l denotes the label, $\mathcal{E}$ the set of labels and U the number of objects.

As mentioned in the introduction, Markov random fields have Gibbs distributions. If the vector ϵ contains the values of the realization of the Markov random field $E(r)$ or $E(p)$, its probability density function $p(\epsilon)$ is given by the Gibbs distribution, see for instance (KOCH AND SCHMIDT 1994, p.258)

$$p(\epsilon) = \frac{1}{Z}\exp(-U(\epsilon))\ , \quad (3.7)$$

where Z denotes the normalization constant and $U(\epsilon)$ the energy. It is determined by summing over the potentials $U_c(\epsilon)$

$$U(\epsilon) = \sum_{c \in C} U_c(\epsilon)\ , \quad (3.8)$$

where c is a clique and C the set of cliques of the graph defined for the Markov random field. We may differentiate between a single-site clique c_1, a two-sites clique c_2, a clique c_q with q nodes up to the clique c_Q with maximum number Q of nodes. Thus we obtain instead of (3.8)

$$U(\epsilon) = \sum_{c_1 \in C_1} U_{c_1}(\epsilon) + \sum_{c_2 \in C_2} U_{c_2}(\epsilon) + \ldots + \sum_{c_q \in C_q} U_{c_q}(\epsilon) + \ldots + \sum_{c_Q \in C_Q} U_{c_Q}(\epsilon) \quad (3.9)$$

with $C_1, C_2, \ldots, C_q, \ldots, C_Q$ being the set of single-site cliques, of two-sites cliques and so on and $C = C_1 \cup C_2 \cup \ldots \cup C_q \cup \ldots \cup C_Q$.

Due to the great number of pixels the normalization constant Z in (3.7) is difficult to compute for image processing. The conditional density $p(\epsilon_i|\partial\epsilon_i)$ of the value ϵ_i of the label at node i with $i \in \{r, p\}$ given the values ϵ_j in the neighborhood, abbreviated by $\partial\epsilon_i$, is therefore used

$$p(\epsilon_i|\partial\epsilon_i) \propto \exp\{-\sum_{c(i) \in C} U_c(\epsilon)\}\ , \quad (3.10)$$

where the sum is taken over the cliques $c(i)$ which contain the node i (KOCH AND SCHMIDT 1994, p.262).

4 Density Functions for the Labeling Processes

At the low level of the image analysis the pixels are labeled according to their affiliation to textures. Clusters of pixels are generally attributed to one texture rather than a few pixels. This fact can be introduced as prior information on the label ϵ_r for the pixel r from (3.2) expressed by the conditional density

$$p(\epsilon_r|\partial\epsilon_r) \propto \exp\{-[\alpha_\epsilon + \sum_{s \in N_r} \beta_s(I(\epsilon_r, \epsilon_{r+s}) + I(\epsilon_r, \epsilon_{r-s}))]\} \quad (4.1)$$

with

$$I(\epsilon_r, \epsilon_q) = \begin{cases} 1\ , & \text{if}\quad \epsilon_r \neq \epsilon_q \\ 0\ , & \text{if}\quad \epsilon_r = \epsilon_q\ . \end{cases}$$

The index s denotes a neighbor of the pixel r in its neighborhood N_r, α_i and β_s the parameters of the density. The parameters α_i control the number of pixels to be attributed to one texture and β_s the directions of the boundaries of the textures. (4.1) is a special case of the Gibbs distribution (3.10) (KOCH AND SCHMIDT 1994, p.313).

The measurements of the gray levels in the different frequency bands of the digital image contain the information on the textures. Let these measurements define a Markov random field so that the density may be obtained by the conditional normal distribution, a special case of the Gibbs distribution. Let $\boldsymbol{y}_r$ be the vector of measurements of gray levels for different frequencies like red, green and blue at the pixel r, the likelihood function then follows with (KOCH AND SCHMIDT 1994, p.308; KOCH 1995)

$$p(\boldsymbol{y}_r|\partial\boldsymbol{y}_r, \epsilon_r, \partial\epsilon_r) \propto \exp\Big\{-\sum_k\{\frac{1}{2\sigma^2_{\epsilon k}}[y_{rk}-$$

$$\mu_{\epsilon k} - \sum_{s\in N_r}\beta_{s\epsilon k}(y_{r+s,k}-\mu_{\epsilon k} + y_{r-s,k}-\mu_{\epsilon k})]^2\}\Big\}, \tag{4.2}$$

where with $\boldsymbol{y}_r = (y_{rk})$ the measurement y_{rk} of the frequency k at pixel r is considered as being independent from the measurement of the different frequencies at pixel r. The parameters of the density are the mean $\mu_{\epsilon k}$ of y_{rk}, its variance $\sigma^2_{\epsilon k}$ and $\beta_{s\epsilon k}$, which describes the texture at the frequency k for identical labels ϵ_r and $\partial\epsilon_r$.

With the prior density (4.1), which can be assumed as being conditionally dependent on $\partial\boldsymbol{y}_r$, and with the likelihood function (4.2) we obtain the posterior density for ϵ_r from (2.3) with

$$p(\epsilon_r|\boldsymbol{y}_r, \partial\boldsymbol{y}_r, \partial\epsilon_r) \propto$$
$$p(\epsilon_r|\partial\epsilon_r)p(\boldsymbol{y}_r|\partial\boldsymbol{y}_r, \epsilon_r, \partial\epsilon_r)\,. \tag{4.3}$$

By means of this density the MAP estimate (2.4) for the label ϵ_r is obtained either by a deterministic or a stochastic procedure (KOCH AND SCHMIDT 1994, p.324). A deterministic method is applied here, which is much faster than a stochastic approach. For each texture training sets are assumed to be available so that the parameters of the density (4.2) can be estimated. Approximate results for the segmentation are used to estimate the parameters of the density (4.1). (4.3) is therefore applied in iterations. The first iteration is solely based on the likelihood function (4.2).

The regions obtained from the segmentation are labeled at the high level of the image analysis in order to find their meaning. Prior information on the label ϵ_p in (3.5) comes from the frequency of the occurrence of the objects in the scene. These frequencies as well as the frequency of the occurrence of neighboring objects are assumed as given by the object model. The prior information is expressed by the conditonal density (3.10) for the label ϵ_p of the region p with $\epsilon = (\epsilon_p)$. By substituting (3.9) we obtain (KOCH 1995; KÖSTER 1995, p.24)

$$p(\epsilon_p|\partial\epsilon_p) \propto \exp\{-\sum_{c_1(p)\in C_1} U_{c_1}(\epsilon) - \ldots$$

$$-\sum_{c_q(p)\in C_q} U_{c_q}(\epsilon) - \ldots - \sum_{c_Q(p)\in C_Q} U_{c_Q}(\epsilon)\}. \tag{4.4}$$

Let S_{pq} be the number of cliques with q sites, thus $S_{p1} = 1$. We then obtain instead of (4.4)

$$p(\epsilon_p|\partial\epsilon_p) \propto \exp\{-U_{p11}(\epsilon) - \ldots$$

$$-\frac{1}{S_{pq}}\sum_{o=1}^{S_{pq}} U_{pqo}(\epsilon) - \ldots - \frac{1}{S_{pQ}}\sum_{o=1}^{S_{pQ}} U_{pQo}(\epsilon)\},$$

$$q \in \{1,\ldots,Q\}\,,\ o \in \{1,\ldots,S_{pq}\}\,, \tag{4.5}$$

where S_{pq} has been used to normalize the contribution of the potentials of the cliques with equal nodes. The potential of one clique is now denoted by $U_{pqo}(\epsilon)$ in order to indicate that the region p belongs to the clique, that q sites constitute the clique and o is its number within the cliques of q sites.

Applying the frequencies mentioned above we may also write

$$p(\epsilon_p|\partial\epsilon_p) \propto p_{p1}(\epsilon)\ldots p_{pq}(\epsilon)\ldots p_{pQ}(\epsilon) \tag{4.6}$$

with

$$p_{pq}(\epsilon) = [\prod_{o=1}^{S_{pq}} p_{pqo}(\epsilon)]^{1/S_{pq}}\,,$$

where $p_{pq}(\epsilon)$ denotes the contribution of the cliques with q sites to the density and $p_{pqo}(\epsilon)$ the contribution of clique o within the cliques of q sites. The number S_{pq} has been used again for the normalization. By comparing the right hand sides of (4.5) and (4.6) we conclude (KÖSTER 1995, p.25)

$$p_{pqo}(\epsilon)^{1/S_{pq}} \propto \exp\{-\frac{1}{S_{pq}}U_{pqo}(\epsilon)\} ,$$

or

$$p_{pqo}(\epsilon) \propto \exp\{-U_{pqo}(\epsilon)\} . \qquad (4.7)$$

Data, which characterize the regions to be interpreted, should be invariant with respect to the scale and the rotation of the regions. Observations describing the form and the compactness of a region are therefore suitable for the interpretation. By forming the ratio of the areas of two regions and the difference of the orientations of two regions, the data for the relations between two regions are obtained. If the ratios and differences are added, relations between three regions are found. This may be continued to higher order relations.

Let the observations for the regions and their relations define a Markov random field. According to the representation (3.9) of the Gibbs distribution, the data for the regions are connected with the one-site cliques and the data for the relations between two and more regions with cliques of two and more nodes. Let $\boldsymbol{y}_p$ be the vector of observations for region p, then

$$\boldsymbol{y}_p = [\boldsymbol{y}'_{p1}, \ldots, \boldsymbol{y}'_{pq}, \ldots, \boldsymbol{y}'_{pQ}]' \qquad (4.8)$$

with

$$\boldsymbol{y}_{pq} = (y_{pqov}) , \qquad q \in \{1, \ldots, Q\} ,$$
$$o \in \{1, \ldots, S_{pq}\} , \; v \in \{1, \ldots, V_q\} ,$$

where $\boldsymbol{y}_{pq}$ denotes the vector of observations for the cliques with q sites and y_{pqov} the observation itself. For each clique with q sites V_q observations are available.

By assuming the components of $\boldsymbol{y}_p$ as being normally distributed and independent although ratios and difference of observations have been used, the likelihood function is obtained by (KOCH 1995; KÖSTER 1995, p.23)

$$p(\boldsymbol{y}_p|\partial\boldsymbol{y}_p, \epsilon_p, \partial\epsilon_p) \propto \exp\{$$
$$-\sum_{v=1}^{V_1} \frac{1}{2\sigma^2_{11v\epsilon}}(y_{p11v} - \mu_{11v\epsilon})^2 - \ldots$$
$$-\frac{1}{S_{pq}} \sum_{o=1}^{S_{pq}} \sum_{v=1}^{V_q} \frac{1}{2\sigma^2_{qov\epsilon}}(y_{pqov} - \mu_{qov\epsilon})^2 - \ldots$$
$$-\frac{1}{S_{pQ}} \sum_{o=1}^{S_{pq}} \sum_{v=1}^{V_Q} \frac{1}{2\sigma^2_{Qov\epsilon}}(y_{pQov} - \mu_{Qov\epsilon})^2\} . \qquad (4.9)$$

The parameters $\mu_{qov\epsilon}$ and $\sigma^2_{qov\epsilon}$ of the density, the expected value and variance of y_{pqov}, depend on the labels ϵ_p and $\partial\epsilon_p$, they are given by the object model.

The posterior density $p(\epsilon_p|\boldsymbol{y}_p, \partial\boldsymbol{y}_p, \partial\epsilon_p)$ is obtained as (4.3) with (4.5), (4.7) and (4.9) from (2.3)

$$p(\epsilon_p|\boldsymbol{y}_p, \partial\boldsymbol{y}_p, \partial\epsilon_p) \propto$$
$$p(\epsilon_p|\partial\epsilon_p)p(\boldsymbol{y}_p|\partial\boldsymbol{y}_p, \epsilon_p, \partial\epsilon_p) . \qquad (4.10)$$

This density is used for the MAP–estimate of ϵ_p. Again the deterministic procedure is much faster than the stochastic one (KÖSTER 1995, p.61).

5 Results

The interpretation has been applied to a part of the RGB aerial photograph "glandorf", which constitutes a test data set of the ISPRS Working Group III/3 (FRITSCH *et al.* 1994). This digital image contains many shadows so that the grey values of the colors red, green and blue were transformed into the HSI color model (hue, saturation and intensity) (GONZALES AND WOODS 1992, p.234). In the HSI color model the shadows appear only in the intensity. Figure 1 shows the intensity by means of grey values of the part of the image "glandorf". Eight training sets were selected from the digital image for the segmentation. They represent streets, bushes, lawns, meadows and shadows from trees and houses. The remaining three sets show three kinds of textures for the roofs of houses. The results of the first segmentation based on the likelihood function (4.2) are given in Figure 2. Figure 3 shows the final results of the segmentation. It contains 389 regions after applying a median filter which corrects minor misclassifications. As a comparison Figure 4 contains the segmentation of the RGB image. The houses cast the intensive shadows, the contours of the houses are better preserved in Figure 3. This segmentation is used for the interpretation, although quite a number of regions are not correctly identified.

Eight objects had to be identified in the digital image: streets (s), houses (ho), annexes or garages (a), hedges (he), lawns (l), meadows (m), bushes (b), greens within streets (g). Observations have been gathered for one-site, two-sites and three-sites cliques, five for the one-site cliques and three for the two- and three-sites cliques. The five observations for the one-site cliques are measures for the form, the roundness and the compactness of the regions and the first and second HU-invariant. The three observations for the two- and three-sites cliques are the area, the orientation and the mean grey level of the regions.

For a comparison the segmented image of Figure 3 has been visually interpreted. The results are shown in Figure 5, where the identified objects are represented by grey values. Large areas labeled as unknown could not be interpreted at all. This is due to the errors in the segmentation. Figure 6 finally shows the results of the automatic interpretation. About 33 % of the objects are incorrectly labeled in comparison to the visually interpreted image. These results are encouraging for further development.

Acknowledgement: This work has been partially supported by Deutsche Forschungsgemeinschaft (DFG).

References

FRITSCH, D., M. SESTER AND T. SCHENK (1994) Test on image understanding. In: EBNER, H., C. HEIPKE AND K. EDER (Eds.), *Proc ISPRS Comm III Symp, München.* SPIE–The International Society for Optical Engineering, Bellingham, 243–248.

GONZALES, R.C. AND R.E. WOODS (1992) *Digital Image Processing.* Addison–Wesley, Reading.

KOCH, K.R. (1988) *Parameter Estimation and Hypothesis Testing in Linear Models.* Springer, Berlin.

KOCH, K.R. (1990) *Bayesian Inference with Geodetic Applications.* Springer, Berlin.

KOCH, K.R. (1995) Markov random fields for image interpretation. *Z Photogrammetrie und Fernerkundung*, 63:84–90, 147.

KOCH, K.R. AND M. SCHMIDT (1994) *Deterministische und stochastische Signale.* Dümmler, Bonn.

KÖSTER, M. (1995) *Kontextsensitive Bildinterpretation mit Markoff-Zufallsfeldern.* Reihe C. Deutsche Geodätische Kommission, München. In preparation.

MODESTINO, J.W. AND J. ZHANG (1992) A Markov random field model-based approach to image interpretation. *IEEE Trans Pattern Anal Machine Intell*, 14:606–615.

PAN, Y. (1994) *Autoregressive Modelle für die Texturanalyse in digitalen Bildern.* Reihe C, 430. Deutsche Geodätische Kommission, München.

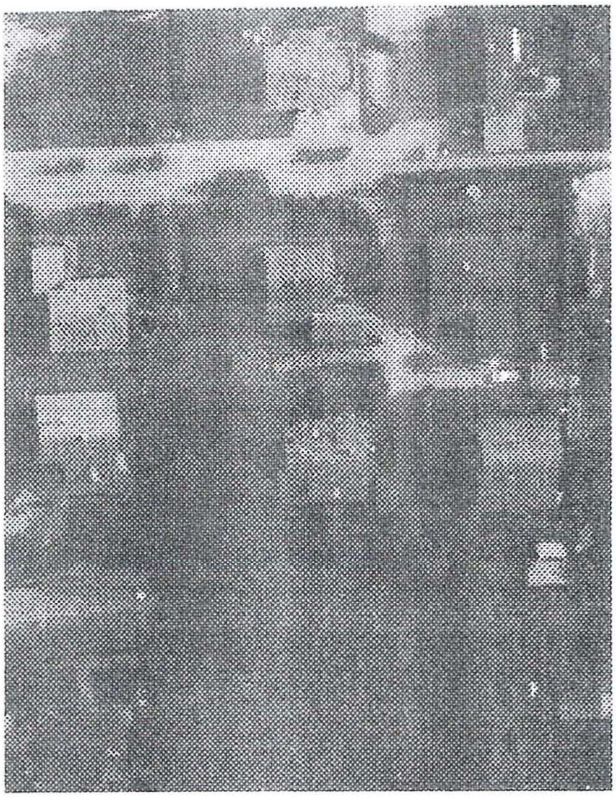

Fig. 1: Intensity of colors of the digital image

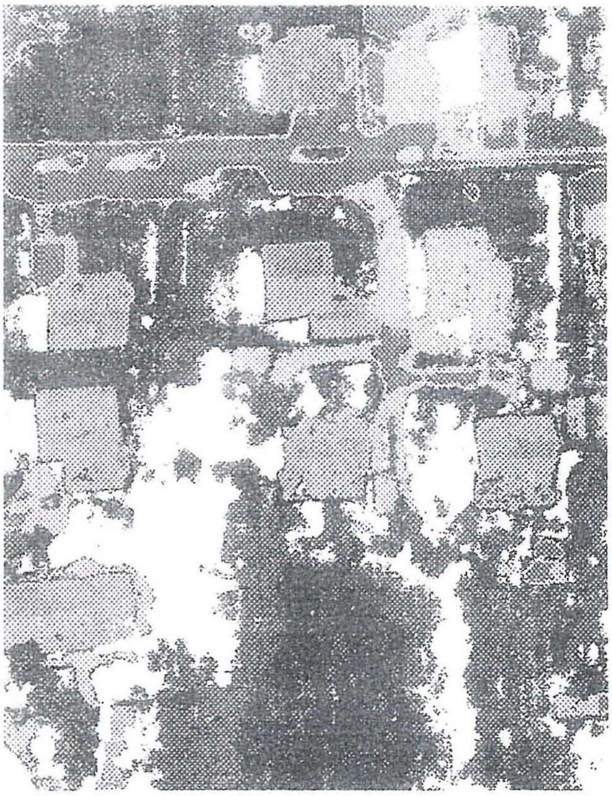

Fig. 2: First segmentation of the HSI image

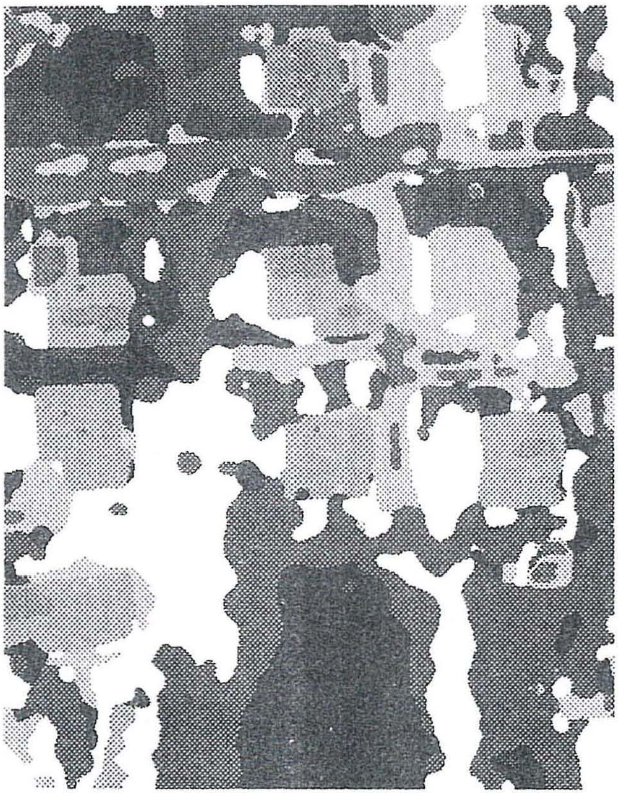

Fig. 3: Segmented HSI image

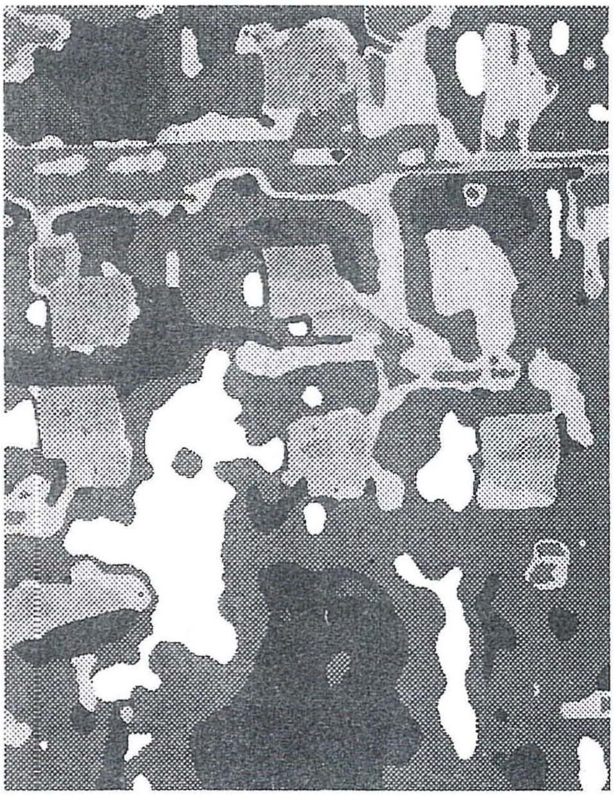

Fig. 4: Segmented RGB image

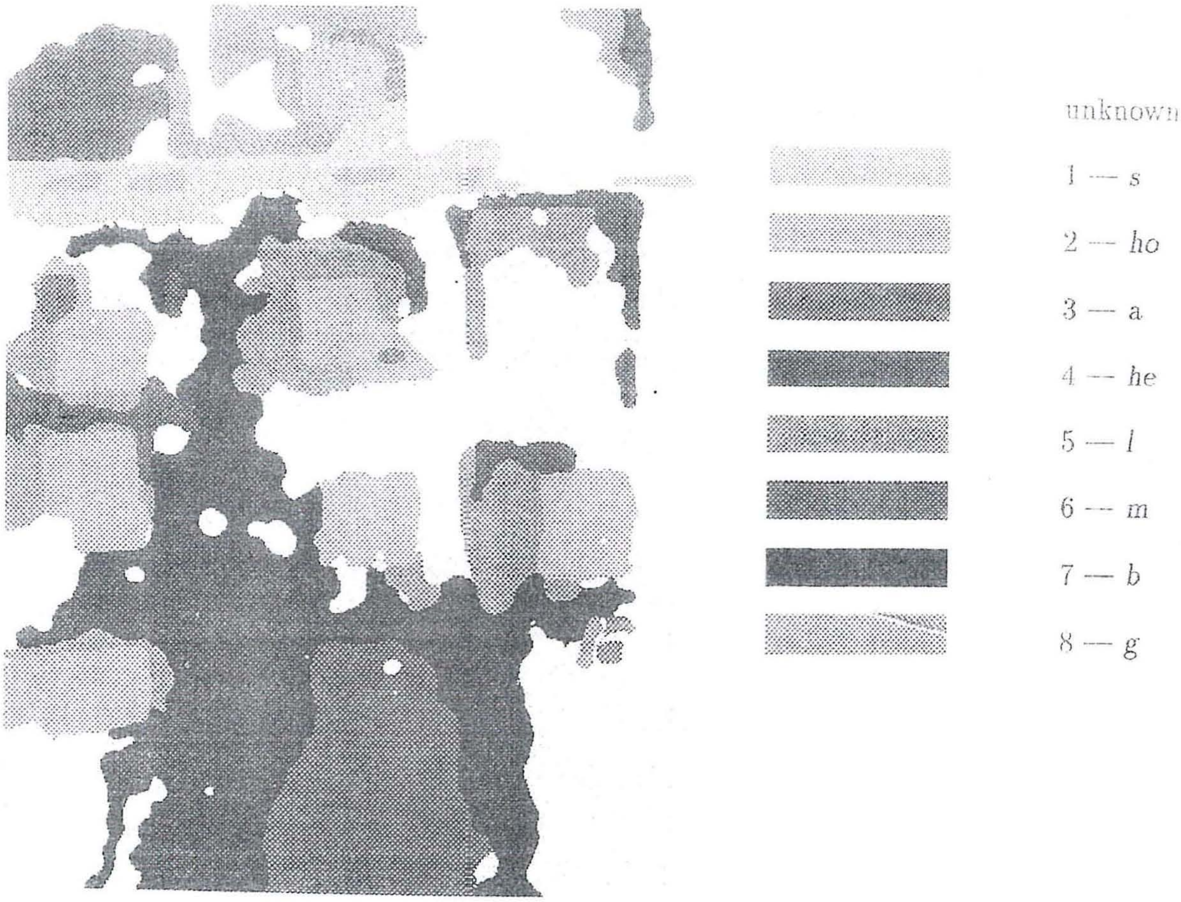

Fig. 5: Visual interpretation

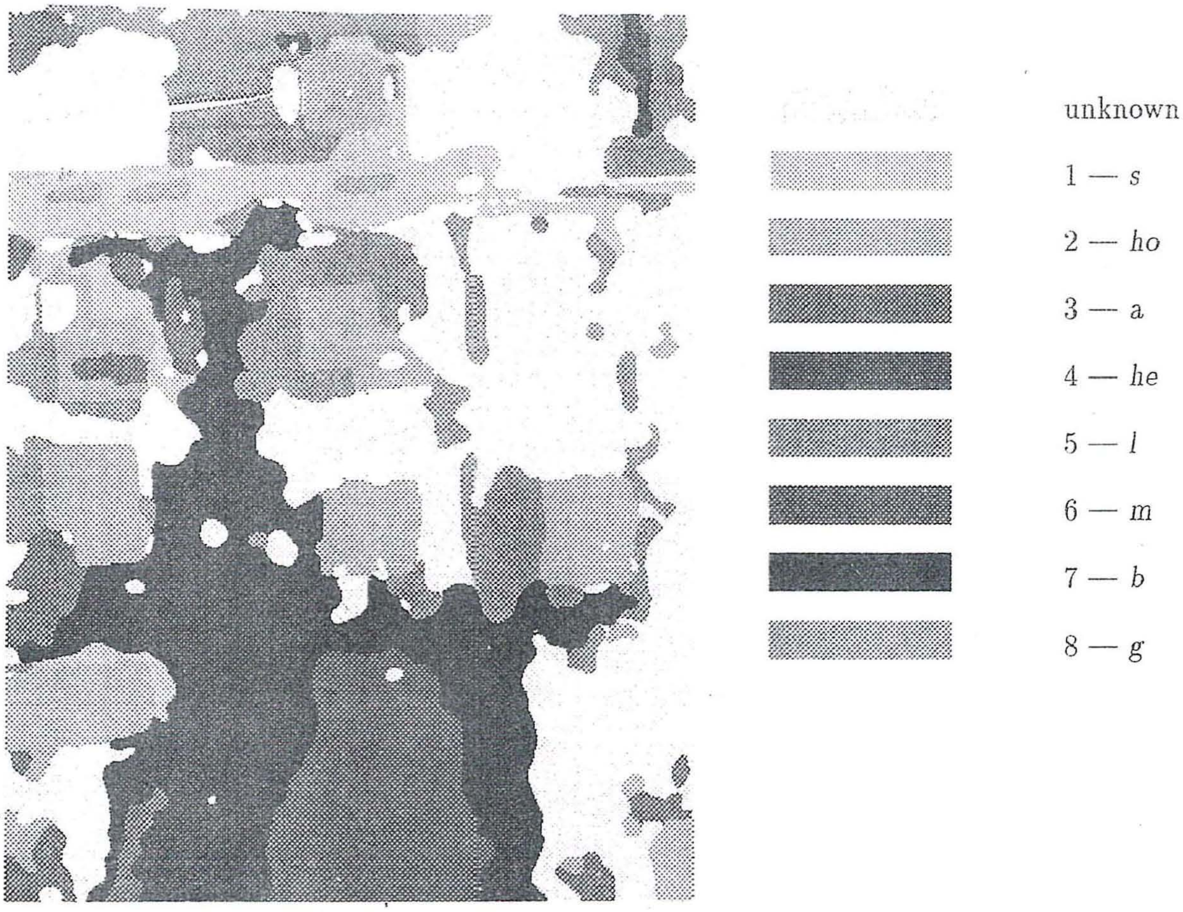

Fig. 6: Automatic interpretation