

Image segmentation based on color and texture gradient

Phu Thien Nguyen

Scientific Center , IBM France
36 Ave Raymond Poincare, Paris 75116

ABSTRACT

An image segmentation scheme based on a model of human perception of color and texture is proposed. It consists of the following steps:

- Spatial segmentation of each RGB images using an edge detector, contour following and closing algorithms;
- Characterization: each intersection of the segments is now characterized by:
 - color: statistical parameters of the RGB values of pixels within the intersection;
 - shape: perimeter, area, compactness;
 - orientation;
 - topology: neighbor, inclusion.
- Merging of the elementary segments based on the selection of their attributes.
- Segmentation by texture gradient based on the window correlation technique to identify "coherent texture" regions.

RESUME

Nous proposons une méthode de segmentation d'image basée sur un modèle de vision de la couleur et de la texture qui consiste en les étapes suivantes:

- La segmentation spatiale des canaux Rouge, Vert et Blue par détection des contours avec un opérateur de gradient, le suivi et la fermeture des contours;
- La caractérisation de chaque intersection des segments par:
 - couleur: paramètres statistiques des valeurs des pixels à l'intérieur d'une même intersection;
 - forme: périmètre, surface, compacité;
 - orientation;
 - topologie: voisinage, inclusion.
- L'agrégation des segment élémentaires par la sélection de leurs attributs.
- La segmentation de l'image par le gradient de texture en utilisant la corrélation des fenêtres sur l'image pour identifier des régions avec des "textures cohérentes".

INTRODUCTION

The understanding of the human visual system and the photo-interpretation method adopted by human analyst does help in designing algorithms in image analysis. The approach is not so much to simulate or worst to imitate the human visual system but rather to understand the underlying mechanisms in order to deduce a few general criteria which could be implemented to extract pertinent features and to identify objects of different visual properties.

Different objects on images can be identified by either labelling pixels of similar properties or defining their boundaries.

The methods of region growing, split and merge, clustering are typical of the first approach. The second approach consists of methods which are based on the detection of contour. Most existing method of contour detection are based on the gradient in the gray level of a half tone image. We are proposing a method of image segmentation which is based on the integrated use of the three fundamental properties: color, texture and geometrical properties of elementary segments on the image. Two methods of contour detection are developed one using a general Sobel operator to detect gradient in gray level and other using a correlation operator to detect texture gradient. These contours are followed to define elementary segments. The elementary segments are then aggregated using a clustering method. Elementary segments are then merged based on their geometrical and topological properties.

The method is applied to the SPOT simulation data taken over an area in the Southwest of Paris.

COLOR PERCEPTION

The trichromatic model of color vision (Faugeras 1976, Pratt 1978, Caelli 1981) is based on the differential absorption spectra of the three pigments found in the cones of the retina. The energy absorbed by the three types of cone are:

$$\begin{aligned} L(x,y) &= \int I(x,y, \lambda)l(\lambda)d\lambda \\ M(x,y) &= \int I(x,y, \lambda)m(\lambda)d\lambda \\ S(x,y) &= \int I(x,y, \lambda)s(\lambda)d\lambda \end{aligned}$$

where:

I: intensity of energy at a point (x,y)

$l(\lambda)$
 $m(\lambda)$: spectral absorption curve
 $s(\lambda)$
 λ : wave length (380 to 800 nM)

L, M and S can be correlated with the primary colors Red, Green and Blue by:

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = U \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

where U is given by CIE (Comission Internationale de l'Eclairage) for the case of TV set phosphorous by:

$$U = \begin{matrix} .3098 & .6321 & .5818 \\ .1208 & .7665 & .1127 \\ .0042 & .1550 & .8408 \end{matrix}$$

The response of the cones seems to be non-linear and L, M and S are transformed into L', M' and S' :

$$\begin{aligned} L' (x,y) &= \text{Log} (L(x,y)) \\ M' (x,y) &= \text{Log} (M(x,y)) \\ S' (x,y) &= \text{Log} (S(x,y)) \end{aligned}$$

A number of experiments seem to suggest that in the visual system, these responses are combined and processed as separate achromatic (A) and chromatic information (C1, C2) which can be expressed mathematically as:

$$A = a(\alpha L' + \beta M' + \gamma S')$$

$$C1 = \mu 1(L' - M')$$

$$C2 = \mu 2(L' - S')$$

The coefficients are given in detail by Faugeras (1976).

A is usually associated with the brightness of the image.

SPATIAL VISION

The visual signals received by retinal receptors are then transmitted to higher level visual cells or group of cells in the Lateral Geniculate Nucleus, and in the visual cortex.

The visual information is processed in a hierarchy depending on the spatial and spectral responses of these visual cells in the receptive fields at different levels. The phenomenon is not yet fully understood, but the following findings (Nevatia 1982, Caelli 1982, Pratt 1976, Tsotsos 1982, Zucker 1984, Creteux 1984) give some insights to how this process is done:

- In the retina, the photo-receptors are grouped in ganglion cells organized into groups of receptive fields. The number of receptors in the ganglion cells is smallest near the fovea and increases nearer to the periphery. This explains the highest visual acuity at the fovea. The spatial arrangement of these receptive fields is suggested to be concentric, hexagonal and organized into a number of layers. The lateral inhibition effect between cells in the receptive field explains the perception of contrast from the information given by the ratio of the response of the central cells to peripheral cells. This contrast sensitivity is a function of spatial frequency of the viewed objects.
- Chromatic response is higher at a lower spatial frequency than achromatic response.
- Chromatic response is relative and the well-known chromatic adaptation effect explains the ability to adapt to local condition of lighting and color balance.
- The receptive fields in the cortex are elongated and hence the response to visual signal at this level is dependent on both spatial frequency and orientation.
- Visual information processing is most precise near the fovea, and the analysis of a scene will require a specific pattern of foveal movement. The analysis of this pattern gives valuable information on how different objects in the scene are being identified.

SEGMENTATION FROM COLOR

An algorithm was developed to detect contours having a strong local gradient (Asfar 1981). The process consists of:

- Detection of edges using a generalized Sobel operator,
- Selection of points having a local maximum gradient,

- Search for the nearest neighbors of each point, taking into account the gradient direction,
- Construction of the contours from the neighbor image.

The detected contours can be followed and closed to isolate different elementary segments. Each segment is then labelled and their attributes computed.

The spatial segmentation is done separately for each of the primary colors Red, Green and Blue. From these segmented images an image of intersection is created where each intersection is given the average R, G and B values of all pixels contained within the intersection. The spatial segmentation of the SPOT data (fig 1) is shown in fig 3. The segmentation can also be done on the L, M and S or A, C1 and C2 components. The R,G and B components are chosen since it was proved that no significant difference is observed when using different color feature sets (Ohta, 1985).

SEGMENTATION BY TEXTURE GRADIENT

The segmentation by texture is treated separately and a texture gradient operator is proposed based on the window correlation technique. The analysis is restricted to chromatic component A (fig 5) because the eye is more sensitive to spatial resolution.

At the boundary of two areas of different textures, the correlation between two windows is maximum along the boundary and minimum across. By calculating the correlation of a central window to eight neighboring windows, the direction and the amplitude (difference between the maximum and minimum correlation coefficient) of the "texture gradient" can be derived. From these two parameters contours can be traced and followed to isolate "coherent texture" regions.

CLUSTERING

An unsupervised classification based on the L1 distance (Ramirez 1982):

$$d_j = \sum_{i=1}^{N_{band}} |\mu_j(i) - x(i)|$$

where : $\vec{\mu}_j$ is the vector of the center of cluster j
 $j = 1, 2, \dots, N_c$ is the number of clusters.

Pixels are assigned to the nearest cluster and its center updated according to:

$$\vec{\mu}_k = \frac{n_k \vec{\mu}_k + x}{n_k + 1}$$

where n_k is the population of cluster k

The clusters found can be:

- split if their standard deviation is below a certain threshold.
- merged if their distance:

$$d_{ij} = \left\{ \sum_{k=1}^N \frac{[\mu_i(k) - \mu_j(k)]^2}{\sigma_i(k) \cdot \sigma_j(k)} \right\}^{1/2}$$

is below a certain value.

A clustering result of the original SPOT data is shown in fig 2 as compared with the result obtained with the segmented images (fig 4).

GEOMETRICAL FEATURE EXTRACTION

The segmented image from either the segmentation by color or texture gradient can be stored under Run Length Coding format (Loodts 1985). The following attributes can then be computed:

Perimeter: The perimeter is defined as the number of pixels belonging to the border of each surface. This is done by accumulating the number of pixels of the same surface surrounded by pixels belonging to other surfaces in two consecutive lines.

Area: this is done by accumulating the length values of pixels belonging to the same surface in two consecutive lines

Compactness: this is a measure given by the parameter $\frac{A}{P^2}$ where P is the perimeter and

A the area of each segment.

Topological features: the relation between segments such as neighborhood, inclusion etc. can be determined by considering pixels belonging to their common boundaries.

ELEMENTARY SEGMENT MERGING

Each elementary segments after the segmentation, clustering and feature extraction are now characterized by a set of parameters:

- color: statistical parameters of the RGB values of pixels within the intersection;
- shape: perimeter, area, compactness, linearity;
- orientation;

- topology: neighbor, inclusion,

The merging can be done by applying a decision based on the selection of these features.

An example of elementary segment merging on the texture classified image (fig 6) by the following criterion:

"Small segments (surface < threshold) included in a large segment (surface > threshold) are merged with the large segment if their forms are not compact ($\frac{A}{P^2}$ is small)".

The result of this merging is shown in fig 7.

CONCLUSIONS

The image segmentation as proposed integrates the spectral and spatial attributes of each pixels in the identification of each elementary segments on a color image. Furthermore these elementary segments can be aggregated by an intelligent selection of their attributes. This presents an approach nearer to the method of photo-interpretation, but there still exists a number of drawbacks which can only be achieved by further researches in the following points:

- The segmentation process is based on global features and local environment can not be taken into account to cater for:
 - local chromatic adaptation;
 - size invariance : a fixed window cannot be used over all regions of wide spectral frequency variation;
 - orientation invariance:
- definition of texture gradient by correlation is still a very crude approximation;
- the segmentation of an image into regions of different color texture with consistent pattern of primitives having a similar color still remains to be investigated. A combined use of texture and color information by averaging the R, G and B segmented images by using the texture class map as a mask is presented in fig 8, this image shows the colors of regions of "similar degree of homogeneity".
- the herarchical approach which can locally adapted to the complexity of different regions on the image.

Acknowledgement

The author is most thankful to Mr A Ballut of IAURIF, Paris for providing the SPOT simulation data over the Paris region.

References

1. Asfar L, "A method for contour detection, segmentation, and classification of Landsat images", International geoscience and remote sensing symposium, Washington D.C., June 1981, pp. 298-304.
2. Cretez J P, L Tanimoto, "Perceptual constancy and the multi-layer visual model: position and size invariances", 7th Int conf on Pattern recognition, Montreal, July 30-Aug 2, 1984,, pp. 518-520.
3. Caelli T, "Visual perception" Pergamon Press, Oxford, 1981, pp. 172-183.
4. Faugeras O, "Digital color image processing and psychophysics within the framework of a human visual model", Ph D dissertation, Unive. Utah, Salt Lake City, June 1976.
5. Loodts J and P.T. Nguyen, "The use of Run Length Coding technique in image storage and processing- an application in remote sensing", The 4th Scandinavian Conference on Image Analysis, Trodheim, June, 1985
6. Nevatia R., "Machine perception", Prentice Hall, New Jersey;1982, pp.90-99.
7. Ohta Y., "Knowledge-based interpretation of outdoor natural scenes", Pitman, Boston, 1985.
8. Pratt W K, "Digital image processing", John Wiley and Sons, New York, 1978, pp. 43-47.
9. Ramirez F, "Local Operation at Hacienda" IBM-UAM Scientific centre, Madrid, Feb 1982.
10. Tsotsos J K, "Knowledge of the visual process: content, form and use", 6th Int Conf on pattern recognition, Munich 1982, pp. 654-668.
11. Zucker S W, A Humel, "Receptive fields and the representation of visual information", 7th Int conf on Pattern recognition, Montreal, July 30- Aug 2, 1984,, pp. 515-517.

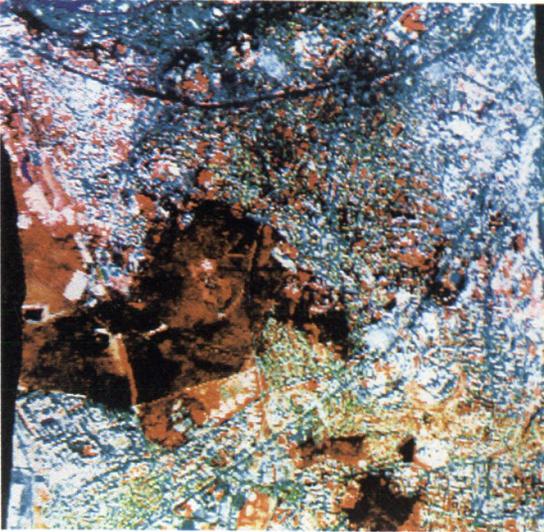


Fig. 1. Color image of the Red, Green and Blue components.

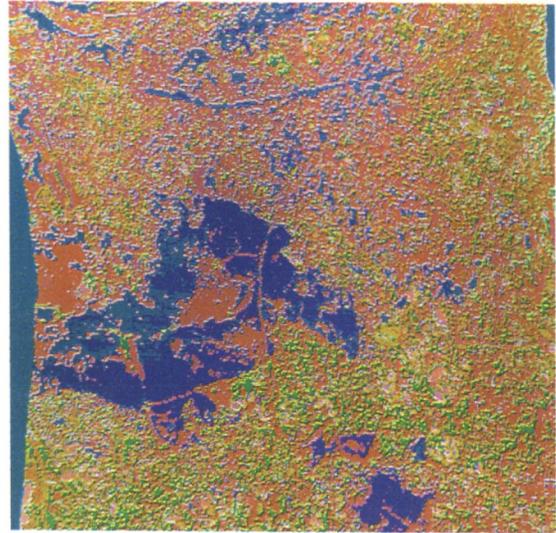


Fig. 2. Classified image of fig 1 by clustering.

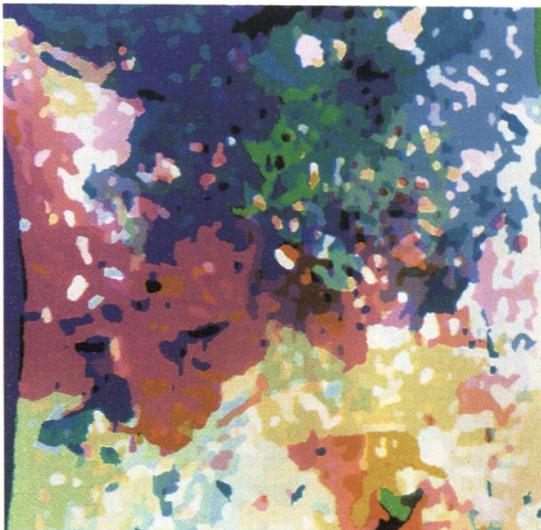


Fig. 3. Color image of the intersection of the three segmented images.

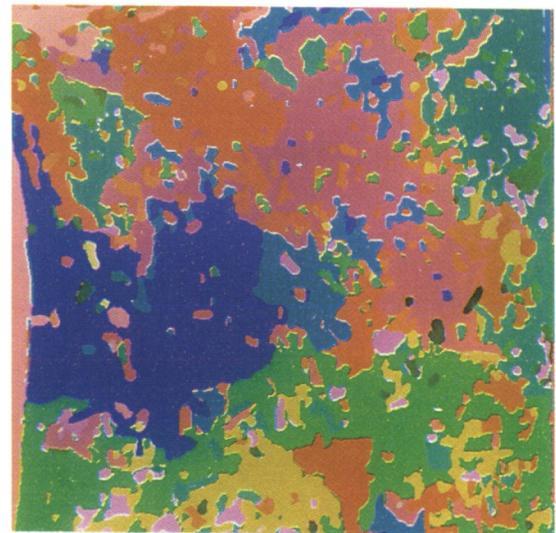


Fig. 4. Classified image of fig 3 by clustering.

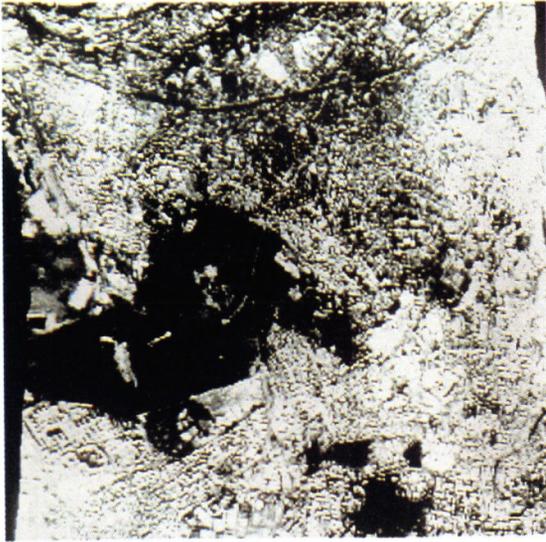


Fig. 5. Achromatic component of the Red, Green and Blue images.

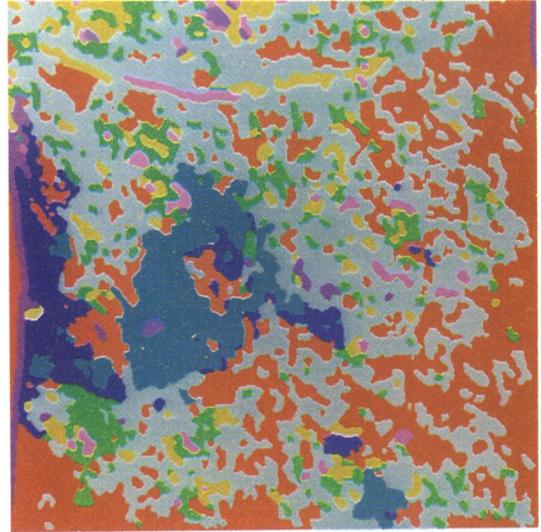


Fig. 6. Classified image of the segmented image by texture gradient after clustering.

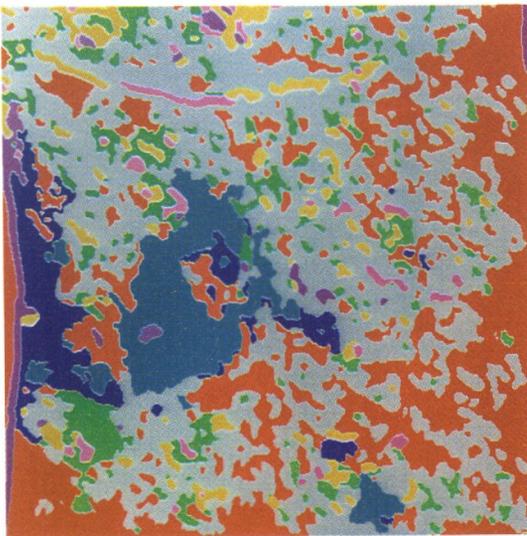


Fig. 7. The result of elementary segment merging of fig 6.

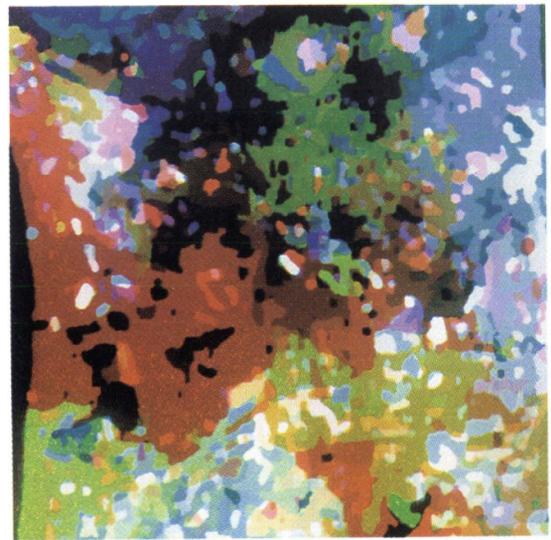


Fig. 8. The result of selective smoothing of fig 3 using fig 7 as a mask.