

INFERENCEAL REGION EXTRACTION IN TV-SEQUENCES

Bernd Radig

Fachbereich Informatik, Univ. Hamburg
Schliiterstrasse 70, 2000 Hamburg 13, W.Germ.

ABSTRACT

In order to speed up the segmentation of image sequences containing moving objects, an approach is proposed which employs knowledge from different sources to control the extraction of regions. Here, regions are connected image components which exhibit grayvalue characteristics determined by specific values of the parameters supplied to the extraction algorithm. These values may be inferred from outside knowledge and previous interpretation results, e.g. from older frames in a sequence. Various segmentations of the same image with parameters adapted to different expected object surface characteristics may be obtained. Thus, this approach is an attempt to overcome the sharp partition of segmentation and interpretation found in conventional scene analysis paradigms.

I INTRODUCTION

Conceptually, image segmentation is a step like any other in computing a hierarchy of abstractions which leads from the pixel representation to the interpretation goal of an image or a sequence of images. "Thus starting from a representation level, an identification process allows to access to an interpretation level, which then becomes the new representation level" [7, p. 8]. Despite the special nature of low-level operators which have to handle a vast amount of data recent research efforts aim at integrating segmentation smoothly into systems which are devised according to methods from the repertoire of artificial intelligence. Typically, in such systems knowledge is formalized and represented explicitly. Examples are: probabilities concerning the interpretation of regions and pairs of neighboring regions [1], constraints on the relationship between objects [8], or relational knowledge databases [4].

One design goal of the region extraction approach, presented here, was to ease the interfacing of low-level region extraction with various knowledge sources. The second design goal concerns the computational efficiency which is particularly important in systems which process image sequences. Since Kelly [3] demonstrated the efficiency improvement of edge detection by "planning", his ideas were carried over to image sequence analysis [10].

Our approach facilitates the controlled focussing on interesting image areas and the fast suppressing of negligible (for the interpretation goal) image components. As this approach was developed for tracing moving objects through TV-sequences it has to deliver, as a third design goal, a region description which is rich and robust enough to permit the computation of reliable similarity values between regions originating from the same natural object projected into different frames. This reliability was demonstrated in [5,6].

II REGION EXTRACTION

The gray tone surface of an 512×573 TV-image is approximated by 256×191 small planes (sloped facets, [2]), obtained from a least-squares fit over 4×5 sized overlapping windows. The azimuth angle (a) of the plane normal is an estimate of the direction of the local image gradient, its polar distance (p) increases with the gradient magnitude, the third coefficient of the plane equation is the mean grayvalue (g). In this way, the original image is transformed into a vector field; regions are defined as connected subsets where plane coefficients exhibit some systematic variation. Therefore, a decision function which labels a vector $v_i = (x_i, y_i, g_i, a_i, p_i)$ as belonging to a region should depend on the vector v_i , its eight neighbors, and a record of parameters (P_1, \dots, P_N) which reflects the desired region characteristic. The choice of parameters is strongly connected with the region descriptor

which consists of frequency distributions H , recorded from the components of vectors in the same region:

H_x, H_y : coordinate values,

H_g, H_a, H_p : plane coefficients,

H_n : number of neighbor pairs;

for two neighbors in the same region:

$H\Delta_g, H\Delta_a, H\Delta_p$: coefficient differences,

$H\Delta_j$: differences of azimuth angles and the angle of the joining line ($\arctan(\Delta y, \Delta x)$).

As bookkeeping information: region number, frame number, and parameter type are kept, the latter identifies the applied parameter record.

The parameters of the decision function mask specified sections of the distributions which must not be populated; they exist as $\{P_x, P_y, P_g, P_a, P_p, P\Delta_g, P\Delta_a, P\Delta_p, P\Delta_j\}$ for the indicated distributions. An example of a region description - marked as A1 in Fig. 1 - is given in Fig. 2. The shaded areas in the histograms mark the forbidden sections. Here, the parameters selected a region of light grayvalue, moderate slope, and not-antiparallel gradients.

This region extraction process meets the design goals.

- It offers a region description which is well suited for computation of similarity values by evaluating the cross correlation of corresponding histograms.
- It is fast to compute. In order to decide if a pair of vectors are in the same region their component values and differences have to be tested against the parameters by a simple boolean operation.
- The parameters have a simple-to-understand meaning. Furthermore, they correspond in such a way to a region description that a description can be easily converted into a parameter record.

III A SIMPLE MOTION EXPERT

This last feature makes the region extraction process especially attractive for use with a knowledge database. Two basic relations exist between regions from an image sequence. As an object projection is usually covered by more than one region those regions form an "intraframe complex" which belong to the same object. These regions move from frame to frame when the object is moving. Each region is connected to a corresponding region in the next frame by interframe links, forming an "interframe sequence". An object can be identified and traced through a sequence of

images as explained for the example of Fig. 3 where a group of pedestrians crosses a street; we are interested in the person with a white shirt and dark trousers. He is modelled as a homogeneous region (T) of dark grayvalue below a region (S) with light grayvalue and an inhomogeneous middle gray region (H) above the shirt. The region T should be 1 to 4 times larger than S. A parameter record PS for S is created (manually, in the current state of implementation). The region extraction process finds some S-regions in the first frame. Depending on their area and position (specified by the bounding rectangle) a parameter record PT is created for each S-region which is used to look for a T-region below. Likewise, a parameter record PH is created to extract the head region above the shirt. A successful triple of H-, S- and T-regions is regarded as a person. All region descriptions are converted to parameter records PH', PS* and PT' which are applied to the next frame. Margins, of course, are added for expected variations. This process continues to the last frame. Fig. 3 shows the traces of the S- and T-centroids.

The whole process can be understood as a simple production system which interacts with a long-term memory containing stored models for objects, and a short-term memory containing regions, intraframe complexes, and interframe sequences. The left hand sides of the production rules are triggered by the contents of the short-term memory. Objects and movements are hypothesized, the right hand sides generate parameter records in order to confirm or falsify these hypotheses by region extraction. Such an approach might fit nicely into an expert system for temporal event recognition [9].

The current implementation of the system contains rules which enable the system to trace rigid objects whose motion is dominated by translation. It can trace an incomplete intraframe complex, so that temporal occlusion of an object part is not disastrous, and it is able to connect broken interframe sequences as long as the regions at the loose ends remain sufficiently similar (white gaps in Fig. 3).

Further work should add rules to the repertoire which allow for more complicated movements, also of non-rigid bodies. Furthermore, situations have to be analyzed which cause sequences to terminate unexpectedly, e.g. due to an occluding object. As a further improvement the user will be able to interactively fill the long-term memory with stored models by presenting examples. Parameter records and relations between regions in an intraframe complex will be automatically

computed from the indicated image components.

REFERENCES

[1] J. A. Feldman, Y. Yakimovsky: "Decision Theory and Artificial Intelligence. 1. A Semantics Based Region Analyzer", Artificial Intelligence 5 (1974) 349-371.

[2] R. M. Haralick: "Edge and Region Analysis for Digital Image Data", Computer Graphics and Image Processing 12 (1980) 60-79.

[3] M. D. Kelly: "Edge Detection in Pictures by Computer Using Planning", Machine Intelligence VI, (1971) 397-409.

[4] M. D. Levine: "A Knowledge-Based Computer Vision System", in: Hanson & Riseman (eds.), Computer Vision Systems, Academic Press, New York 1978.

[5] B. Radig: "Description of Moving Objects Based on Parameterized Region Extraction", in Proc. IJCP-78, Kyoto, Nov. 1978, 723-725.

[6] B. Radig: "Image Region Extraction of Moving Objects", to appear in: T. S. Huang (ed.), Image Sequence Analysis, Springer, Berlin-Heidelberg-New York, 1981.

[7] J. C. Simon, E. Backer, J. Sallantin: "A Structural Approach of Pattern Recognition", Signal Processing 2 (1980) 5-22.

[8] J. M. Tenenbaum, H. G. Barrow: "Experiments in Interpretation Guided Segmentation", Artificial Intelligence 8 (1977) 241-274.

[9] J. K. Tsotsos: "Temporal Event Recognition: An Application to Left Ventricular Performance", In Proc. IJCAI-81, Vancouver, August 1981.

[10] M. Yachida, M. Ikeda, S. Tsuji: "A Plan-Guided Analysis of Cineangiograms for Measurement of Dynamic Behaviour of Heart Wall", IEEE Trans. Pattern Analysis and Machine Intelligence PAMI-2 (1980) 537-543.

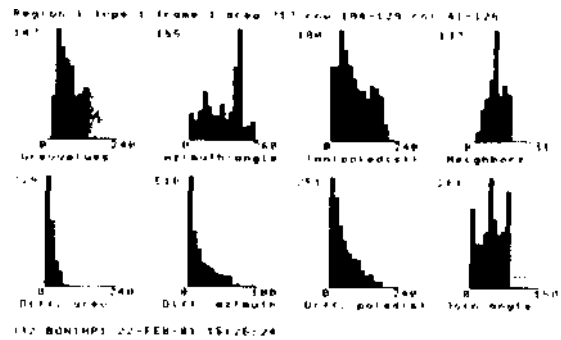


Fig. 1 Attributes of region A1



Fig. 2 Image sequence of moving car

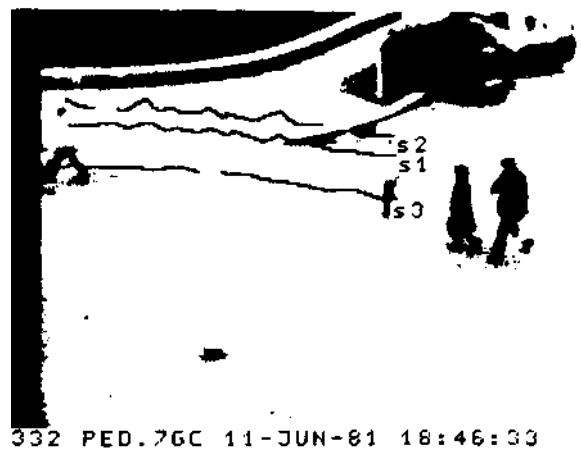


Fig. 3 Moving intraframe complex: one pedestrian selected