

AUTOMATIC DETECTION OF INTEREST AREAS OF AN IMAGE OR OF A SEQUENCE OF IMAGES

X. Marichal, T. Delmot, C. De Vleeschouwer, V. Warscotte and B. Macq

*Laboratoire de Télécommunications et Télédétection
Université catholique de Louvain
2, place du Levant - B 1348 Louvain-la-Neuve - Belgium
Phone: +32 10 47 41 05 - Fax: +32 10 47 20 89
E-Mail: marichal@tele.ucl.ac.be*

ABSTRACT

The tool introduced in this paper allows to automatically decide in an image or in a video sequence which regions are important and which ones are not. For this purpose, *Fuzzy Logic* has been used to modelize human subjective knowledge about the way to allocate priorities to regions. The resulting classification can be used in a wide range of applications going from image coding to image understanding.

1. INTRODUCTION

Nowadays, most of image processing applications require a preliminary analysis and comprehension of the pictures contents. Conceptually, there are two major approaches for image understanding. On the one hand, *image-based* methods detect and segment image regions that correspond to real objects or object parts. They then compare these objects with memorized objects that are present in the solution domain (e.g. using pattern recognition techniques). These methods are called *bottom-up* methods because they proceed from the raster image to describe it by regions. On the other hand, *model-based* methods generate and test a set of assumptions from applicable knowledge in order to decide whether or not the image verifies the model. Two strategies can be distinguished: the hierarchical one or the non hierarchical one. The hierarchical strategy (also called *top-down* strategy) verifies each hypothesis in a fixed order while the non-hierarchical one verifies each hypothesis in a random order, each partial result contributing to the elaboration of the final one. It can be understood as a cooperation of competitive experts. Nevertheless, it quickly appeared that the best methods were the ones combining both philosophies. In particular, the tool presented here follows a combined approach. Beginning with the previously explained segmentation of the original picture (*bottom-up* approach), a non-hierarchical fuzzy model allows to combine the

results of distinct criteria in order to increase the reliability of the system and to provide convergence to a meaningful final result. The use of Fuzzy Logic System (FLS) methodology allows to process semantic inference through a mathematical formalism. This ability to treat semantic statements is especially useful to deal with the complexity of psychological factors involved in image understanding.

2. IMAGE SEGMENTATION

Before studying their interest, extraction of picture regions is requested. In the present framework, segmentation aims at dividing the picture in regions corresponding to real objects or parts of real objects in the scene. For the following analysis purposes, the extracted areas should satisfy two main constraints. First, picture elements assigned to a region have to be related to a single real object. Regions have thus to follow the real objects edges. Secondly, the regions should be as large as possible because the larger the amount of data taken into account for the posterior analysis of interest, the more this analysis will be efficient. Practically, regions of uniform spatial features will be extracted in two steps:

First, a watershed algorithm based on immersion simulations[1] divides the image in small regions whose edges include real objects edges. The gradient flooded by the immersion process is a weighted sum of the gradients evaluated on each picture component (luminance + chrominance). Oversegmentation inherent to this approach can be reduced by thresholding the gradient used for the immersion process and by merging, during the flooding step, the neighboring regions having areas lower than a particular threshold. Nevertheless, the results obtained by this method show that these thresholds have to be kept at a rather low value if one wants to avoid merging parts of distinct real objects in a unique region.

The second step tries to reduce the unavoidable

oversegmentation due to the first step. It consists in merging adjacent regions having close spatial features. In practice, regions having close luminance and chrominance mean values are merged.

3. DETERMINATION OF THE LEVEL OF INTEREST OF THE REGIONS

The aim of this section aims at describing a tool able to attribute a level of interest to the different areas of a picture. This will be made possible through the synthetization of a FLS (readers not familiar with the terms and concepts of Fuzzy Logic may refer to [2] and [3]).

3.1. Subjective Knowledge: Interest Criteria

When dealing with semantic interpretation of images, one unique criterion appears to be insufficient. It is obvious that the human brain, thanks to its *a priori* knowledge and its huge memory containing real-world concrete scenes, combines different subjective criteria in order to assess its final decision. Up to now, five of these criteria have been distinguished. Three of them are designed for intra-images while the two others take into account the temporal component particular to image sequences.

The **interactive criterion** eliminates regions with no pixels inside a pre-defined square window of the image. This criterion is the interactive criterion *par excellence* as the user can easily select another "interest window". The **image border criterion** is based on the fact that the most important part of an image in a video sequence is correctly centered and that the eye precision is better in the center of the vision area than in its extremities; it rejects the regions which have a huge amount of their pixels along the picture borders. The **face texture criterion** is mainly designed for videophone or video-conference; it rejects all the regions whose texture components do not coincide with a set of skin samples.

A **motion criterion**, based on a succession of motion estimations between the successive images, reduces the interest of regions with no motion and a very small gradient (the Human Visual System do not focus on these areas) as well as the regions with a very important motion and a high gradient (they are too noisy to be correctly perceived). The **continuity criterion** takes into account the result obtained for the previous image in order to give some temporal stability to the final evaluation of the level of interest.

3.2. Knowledge Formalization: Fuzzy Rules and Fuzzy Sets Definition

Each criterion can be expressed in terms of fuzzy rules manipulating linguistic variables. The consequence of each rule will concern the same solution variable: the "INTEREST LEVEL".

For the **interactive criterion**, the input variable of the rules is the percentage of pixels of the region belonging to the coarsely predefined interest area. This variable is noted "HIT PERCENTAGE". According to the criterion understanding, the following rules can be deduced: *If hit percentage is HIGH then interest is VERY HIGH* and *If hit percentage is LOW then interest is LOW*. The second rule, which is the negation of the first one, is necessary in order to reduce the interest of a region that is outside the predefined interest area.

For the **image border criterion**, the input variable handles the fraction of pixels of the region frontier that belongs to the image border. Hereunder, this variable is noted "FRACTION FRONTIER PIXELS". The rule expressing the criterion is: *If the fraction frontier pixels is HIGH then interest is LOW*.

For the **face texture criterion**, the input variable is the geodesic distance [1] between a kernel of points corresponding to face texture samples and the point representing the region in a chrominance space. This space is a two dimensional space in which each region is represented by its mean chrominance values. The handled input variable is named "FACE TEXTURE DISTANCE" and the rule expressing the criterion is: *If face texture distance is VERY LOW then interest will be HIGH*. In a sequence of videophone or video-conference, it may be interesting to go further by reducing the interest of areas whose texture are not face texture. So, a second rule could be added: *If face texture distance is HIGH then interest is (SOMEWHAT) LOW*.

For the **motion criterion**, there are two input variables. The first one is a measure of the motion of the studied area. The corresponding linguistic variable is denoted "MOTION". This motion value is normalized by the maximum motion value (fixed by the search window size that has been chosen for the motion estimation). It is also interesting to note that the motion measure uncertainty could be considered thanks to the fuzzy numbers. The second variable is a measure of the area contrast. The linguistic variable "CONTRAST" denotes this notion. The used value has been a mean value of the picture gradient. Having the input variables defined, three rules express the motion criterion: *If motion is SMALL and contrast is SMALL then interest is LOW*, *If motion is (VERY) HIGH and contrast is (SOMEWHAT) HIGH then interest is LOW* and *If*

motion is *MEDIUM* then interest is (*SOMEWHAT*) *HIGH*.

The continuity criterion aims at keeping a constant interest level for an object during the whole sequence. Of course, the correspondence between the regions detected at two different times would require a tracking procedure and the level of interest of a region of the first image can be projected to the second image only if the tracking of the region is reliable enough. The linguistic variable involved in this criterion is the "TRACKING RELIABILITY" and the rule can be expressed as: *If tracking reliability is HIGH then interest is AROUND previous value*. The previous value corresponds to the interest value of the region at time $t - 1$ associated to the considered region at time t . For the first tests, the tracking has been based on the improbable hypothesis that there was no motion. The correspondence between regions of pictures at time t and $t - 1$ is thus immediate and the evaluation of its reliability is based on the computation of the pixels by pixels difference between the two pictures.

Despite the fact that the rules can be clearly formulated, they still contain a large amount of subjectivity in their interpretation. This subjectivity will be formalized through the following enumeration of the involved linguistic variables and the attached fuzzy sets definitions:

1. **"INTEREST"** The universe of discourse of the variable is $[0,1]$. Each term of this variable is a fuzzy set whose membership function is defined on the universe of discourse. The terms are *HIGH*, *LOW* and *AROUND* (a given value). Their membership functions are triangular and illustrated on figures 1 (a.1) and (a.2).
2. **"HIT PERCENTAGE"** The universe of discourse of the variable is $[0,1]$ (it is a percentage). The terms are *HIGH* and *LOW*. They are defined by membership functions of figure 1 (b).
3. **"FRACTION FRONTIER PIXELS"** The universe of discourse of the variable is $[0,1]$. The only term is *HIGH* and it is also defined by a triangular shape function (figure 1 (c)).
4. **"FACE TEXTURE DISTANCE"** The distance used has been normalized by its maximal value. So, its universe of discourse is also $[0,1]$. The terms are *LOW* and *HIGH*. Figure 1 (d) illustrates them.
5. **"CONTRAST"** The mean value used as the contrast value is normalized by its maximal value (255). So, once again, the universe of discourse is $[0,1]$ and the terms *HIGH* and *LOW* are defined by the triangular shapes of figure 1 (e).

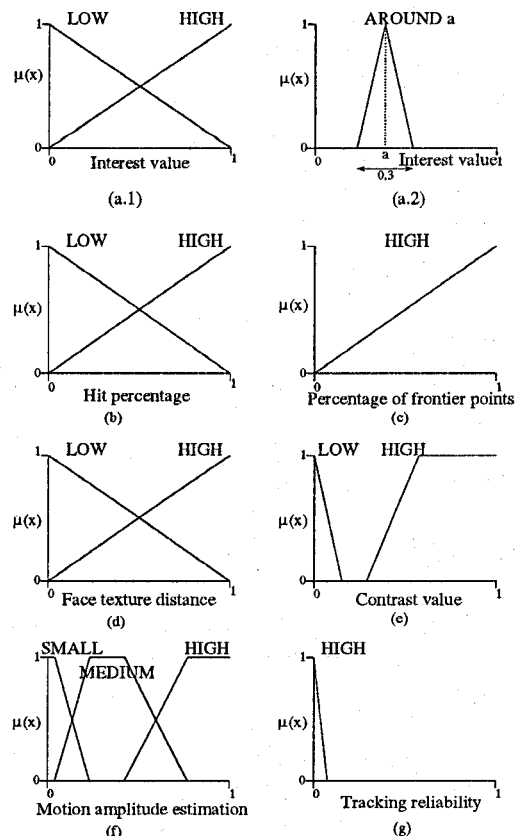


Figure 1: Representation of the membership functions defining fuzzy sets for each used linguistic variable.

6. **"MOTION"** The universe of discourse is $[0,1]$. The three terms are *SMALL*, *MEDIUM* and *HIGH*. They are defined by the trapezoidal or triangular membership functions drawn on figure 1 (f).
7. **"TRACKING RELIABILITY"** The summation of the pixel by pixel difference is divided by the number of pixels and by the maximal possible difference. So, the universe of discourse of the tracking reliability variable is also $[0,1]$. The term *HIGH* is represented on figure 1 (g).

3.2.1. FLS Implementation

In order to implement a tool, one has still to fix the *t-norm* and *t-conorm*, the implications rules, the aggregation method and the defuzzifier.

In the first implementation, "*max*" and "*min*" have been chosen as the *t-norm* and *t-conorm* operators while the *minimum implication* was chosen for fuzzy reasoning. For the aggregation process, the *additive aggregation* technique has been used and, for computation

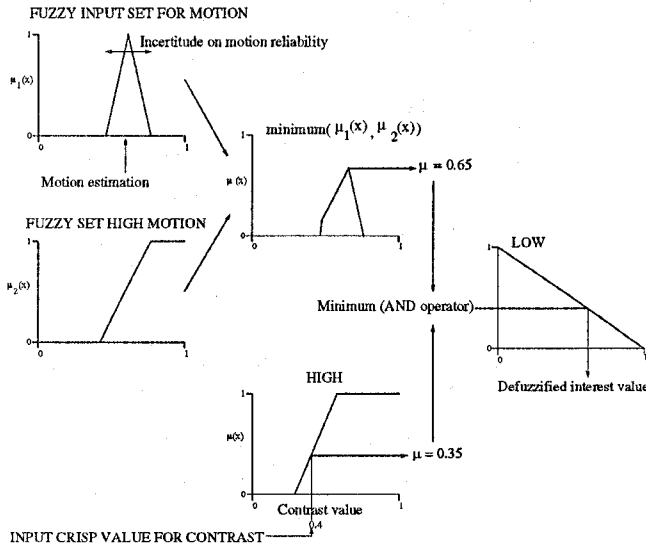


Figure 2: Example of the way rules are working. Motion rule: "If motion is HIGH and contrast is HIGH then interest is LOW".

simplification purpose, the *height defuzzifier* was chosen to combine the defuzzified values of each rule output subset. Such a defuzzification procedure permits to easily take into account the confidence that the designer has about each rule, by allocating an adequate weight to each of them. The defuzzified value of each set is provided by the *maximum defuzzifier*. In case of several maximum, the left, right or median one is chosen if the fuzzy membership function is respectively monotonic increasing, monotonic decreasing or non-monotonic. Figure 2 illustrates how a rule is treated.

As a conclusion, with w_i the weight of each rule R^i , B^i the output fuzzy set of rule R^i and $\bar{y}^i$ the value defuzzified from B^i , the final result $\bar{y}$ is computed by:

$$\bar{y} = \frac{\sum_{i=1}^M w_i \mu_{B^i}(\bar{y}^i) \bar{y}^i}{\sum_{i=1}^M w_i \mu_{B^i}(\bar{y}^i)} \quad (1)$$

4. CONCLUSION

First results are promising. Nevertheless, rather large improvements can be brought to the process tested as a first implementation. A first improvement would consist in the refinement and completion of the criteria. As explained before, the motion and continuity criteria could be improved by an evaluation of the motion estimation reliability. The criteria completion by a user having specific field of interests in the image he/she is watching at is obvious: the flexibility in the choice of the criteria and the weighting of each of them is a ma-

jor feature of the tool that enables to easily take into account additional criteria.

Another improvement to the final result could be brought by taking into account the neighboring relationships between regions. This would permit to have a spatially more coherent distribution of the levels of interest. This problem is obviously related to the segmentation one. In other words, research towards a more effective segmentation including both spatial filters and tracking information would permit to have spatially and temporally more coherent decision.

Hereabove, the FLS rules have been chosen according the subjective knowledge our brain manipulates to interpret the picture features. Nevertheless, having identified some factors influencing the importance of a region in the cognitive information system brought by the pictures, it could be interesting to identify FLS rules in a more objective way. Clustering methods for instance allow such identification from couples of (*input*, *output*) vectors [4]. The definition of such couples requires to be able to attribute a reliable level of interest to the regions of a training picture. Some research based on the ocular motion measurement have already produced results [5] and would probably be an interesting way to objectively match a level of interest to the training data.

5. REFERENCES

- [1] L. Vincent and P. Soille. Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. *IEEE Trans. on Pat. Anal. and Mach. Int.*, 13(6):583-598, June 1991.
- [2] Earl Cox. *The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems*. Academic Press Limited, 955 Massachusetts Avenue, Cambridge, MA 02139, 1994. ISBN 0-12-194270-8.
- [3] Jerry M. Mendel. Fuzzy Logic Systems for Engineering: A Tutorial. *Proceedings of IEEE*, 83(3):345-377, March 1995.
- [4] Bart Kosko. Fuzzy Associative Memories. In *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, chapter 8, pages 299-337. Prentice-Hall International, Inc., University of Southern California, 1992. ISBN 0-13-612334-1.
- [5] A. Monot, B. Drobe, F. Mekhazni, P. Denieul, P. Viviani, and T. Alpert. Analyse comparative des mouvements oculaires: Application à la mesure objective de la qualité d'image. pages 187-195, Grenoble, February 1996. CORESA 96.