A Fuzzy Logic System Able to Detect Interesting Areas of a Video Sequence

C. De Vleeschouwer*, X. Marichal†, T. Delmot‡and B. Macq

Laboratoire de Télécommunications et Télédétection
Université catholique de Louvain
Bâtiment Stévin - 2, place du Levant
B-1348 Louvain-la-Neuve
Tel.: +32 10 47.41.05 - Fax: +32 10 47.20.89
E-mail: devlees@tele.ucl.ac.be

ABSTRACT

This paper introduces an automatic tool able to analyse the picture according to the semantic interest an observer attributes to its content. Its aim is to give a "level of interest" to the distinct areas of the picture extracted by any segmentation tool. For the purpose of dealing with semantic interpretation of images, a single criterion is clearly insufficient because the human brain, due to its a priori knowledge and its huge memory of real-world concrete scenes, combines different subjective criteria in order to assess its final decision. The developed method permits such combination through a model using assumptions to express some general subjective criteria. Fuzzy Logic enables to encode knowledge in a form that is very close to the way experts think about the decision process. This fuzzy modelling is also well suited to represent multiple collaborating or even conflicting experts opinions. Actually, the assumptions are verified through a non-hierarchical strategy that considers them in a random order, each partial result contributing to the final one. Presented results prove that the tool is effective for a wide range of natural pictures. It is versatile and flexible in that it can be used stand-alone or can take into account any a priori knowledge about the scene.

Keywords: Fuzzy Logic System, Video Analysis, Semantic Interpretation, Segmentation, Compression.

1 INTRODUCTION

Nowadays, many image processing applications require a preliminary analysis and comprehension of the content of the pictures. This need is obvious in the case of pattern recognition processes in which picture analysis is in fact the application kernel. For medical imaging or for satellite picture interpretation, semantic analysis also remains a major research area. In most cases, however, the efficiency in analyzing and interpreting the picture is directly related to the use of a model including *a priori* knowledge about the described scene. The greatest difficulty encountered while developing a universal picture analysis tool arises from the unavoidable subjectivity related to the scene observation.

*Research of C. De Vleeschouwer is supported by a grant F.N.R.S. - Alcatel Bell.
†Research of X. Marichal is supported by a grant of the FRIA.
‡Research of T. Delmot is supported by Belgacom.

The goal of this paper is to introduce an automatic tool able to analyze the different areas of a picture and to attribute them a "level of interest". This tool is effective for a very large range of natural pictures. It is versatile and flexible in that it can be used stand-alone or can take into account an *a priori* knowledge about the scene. Conceptually, there are two major approaches to image understanding[1]:

- *Image-based* methods detect and segment image regions that may correspond to real objects or to portions of object, and eventually match them in a database. These methods are also called *bottom-up* methods because they proceed from the raster image to describe it by regions.
- On the other hand, *model-based* methods seek to determine whether or not the image fits 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 assumption in a fixed order while the non-hierarchical one verifies each assumption in a random order, each partial result contributing to the elaboration of the final one. It can be viewed as a cooperation of competitive techniques.

However, the best methods appear to be the ones combining both philosophies. Beginning with the segmentation of the original picture (*bottom-up* approach), a non-hierarchical fuzzy model allows us 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 a Fuzzy Logic System (FLS) methodology allows the processing of semantic inference through a mathematical formalism. Besides, some problems in object recognition have already been solved using FLS[2]: image recognition using c-means clustering,[3,4] recognition of moving objects[5] and recognition of vegetables using linguistic expressions[6] are some interesting examples which demonstrate how fuzzy logic can deal with ambiguity and subjective knowledge. The complete classification algorithm and more details about the use of fuzzy logic are presented in section 2. Section 3 presents some results for the region classification. It is demonstrated that the tool performs correctly even with over segmentation inherent to current available segmentation tools.

2 REGIONS CLASSIFICATION

Preliminary remark:

In this section, strong hypothesis is made that extracted regions represent meaningful objects of the scene. Which means that the segmention process aims at dividing the picture in regions corresponding to real objects or parts of real objects in the scene. Nevertheless, results of section 3 are obtained with a classical segmentation algorithm extracting regions of spatial uniformity, and demonstrate the tool efficiency even when extracted regions do not match real objects of the scene.

A criterion of interest is searched for. It must be able to automatically attribute a level of interest to the various parts of a picture. As the first aim of the tool was to perform "subjective video coding", the semantic interest is actually directly related to the identification of the features of the scene. An area containing lots of meaningful features (these are important features for the semantic understanding of the scene) has to be classified among interesting areas and should be treated carefully through the coding process. the remaining question is: "How can areas with meaningful features be selected?". For the purpose of dealing with semantic interpretation of images, a single criterion is clearly insufficient. Obviously, the human brain, due to its *a priori* knowledge and its huge memory of real-world concrete scenes, combines different subjective criteria in order to assess its final decision. According to this logic, it has been decided to combine several criteria in order to provide the final classification. In addition, the criteria combination has to be robust against noise, which may not be the case if a single criterion is used.

The tool for classification developed here uses a Fuzzy Logic System (FLS,[7,8]). Indeed, *fuzzy sets* are a mathematical methodology that was developed in order to express the semantic ambiguity inherent in human

language. They are an efficient tool to deal analytically with subjectivity. Subjective knowledge, in contradiction with objective knowledge, that can be strictly formulated through mathematical models, is usually impossible to quantify using traditional mathematics. A significant benefit of fuzzy system modeling is the ability to encode knowledge directly in a form that is very close to the way experts themselves think about the decision process; a fuzzy system captures expertise close to the expert's own cognitive model of the problem. Another benefit from FLS is that they are well suited to represent multiple cooperating, collaborating and even conflicting experts opinions.

Moreover, it can be demonstrated[7] that a FLS is a non linear system that maps a crisp input vector into a crisp scalar output. It can also be expressed as a linear combination of *fuzzy basis functions* and, in this sense, can be viewed as a universal function approximator. Working with fuzzy logic does not enforce restrictive constraints to the problem solution domain. It is obvious that, as it consists of combining somewhat ambiguous criteria in order to result in a single crisp output (= level of interest), the problem being considered is well suited to be solved using FLS. The fact that the knowledge about the quality requirements can only be subjectively expressed using human language enforces the relevance of FLS use.

2.1 Application of Fuzzy Logic for Regions Classification

The aim of this section is to describe the way of attributing a level of interest (related to the required quality) to the different areas of a picture. The following steps are followed:

- The first step is to identify the *a priori* knowledge one has about "What is an interesting area?". This knowledge is of course totally subjective.
- The second step is to express the subjective knowledge with fuzzy rules. It will also define the linguistic variables manipulated by these fuzzy rules, defining appropriate membership functions.
- The third and last step involves fixing the t-norm and t-conorm,[9] implications rules,[10] the aggregation method, the fuzzifier and the defuzzifier.[8] These determine how fuzzy reasoning performs.

2.1.1 Subjective Knowledge: Some Interest Criteria

As mentioned earlier, for the purpose of dealing with semantic interpretation of images, a unique criterion appears to be insufficient. Up to now, five subjective criteria have been distinguished. Three of them are designed for intra-images while the two others take into account temporal components specific to moving images.

Intra-images criteria

- **Interactive criterion**: Its principle is to decrease interest of 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".
- **Image border criterion**: It is based on the fact that the most important part of an image in a video sequence is at the center of the image. It allocates low interest values to the regions with significant number of pixels on picture borders.
- **Face texture criterion**: It is mainly designed for video-phone or video-conference. This criterion decreases the interest of regions whose texture components do not coincide with a set of skin samples.

Inter-images criteria

- **Motion and contrast criterion**: It is based on a succession of motion estimations (using[11]) between the images at the highest frequency (even if the sequence is coded at $8Hz$, the criterion will compute three motion estimations between images at $25Hz$), it reduces the interest of regions with no motion and a very small activity (the Human Visual System does 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). Activity and gradient are defined section 2.1.2.A.
- **Continuity criterion**: It takes into account the result obtained with the previous image in order to give some temporal stability to the final evaluation of the level of interest.

2.1.2 Knowledge Formalization: Fuzzy Rules and Fuzzy Sets Definition

A) Rules

Each criterion is expressed in terms of fuzzy rules manipulating linguistic variables. The consequence of each rule relates to the same solution variable: the "INTEREST LEVEL". This variable is redefined into a set of fuzzy regions according to the comprehension of every rule.

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.
- 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 boundary that belongs to the image border. Hereafter, this variable is noted "FRACTION BOUNDARY PIXELS". The rule expressing the criterion is:

- If the fraction border pixels is HIGH then interest is LOW.

For the **face texture criterion**, the input variable, named "SKIN PROBABILITY", is an estimation of the probability $P(SKIN \mid C_b, C_r)$ that the region corresponds to skin samples knowing the region representative point in a chrominance space. A desirable property for the used color coodinate system could be that any unit change in the chromaticity diagram should be perceived by an observer as an equivalently noticeable color shift. The *C.I.E. Uniform Chromaticity Scale Color Coordinate System* ($= (L, u, v)$ system) is an attempt for obtaining such a system.[12] Nevertheless, as our criterion only focuses on a particular area of the chromaticity diagram, the amount of computation to transform the available (Y, Cb, Cr) coordinates into (L, u, v) color coordinates is not justified. So, the used color space is the (Cb, Cr) space and, in this space, each region is represented by its mean chrominance values. Using Bayes's rule, $P(SKIN \mid C_b, C_r)$ is expressed by:

$$P(SKIN \mid C_b, C_r) = \frac{T(C_b, C_r \mid SKIN) \cdot P(SKIN)}{T(C_b, C_r)} \quad (1)$$

$$= \frac{T(C_b, C_r \mid SKIN) \cdot P(SKIN)}{T(C_b, C_r \mid SKIN) \cdot P(SKIN) + T(C_b, C_r \mid \overline{SKIN}) \cdot P(\overline{SKIN})} \quad (2)$$

In this expression, $P(SKIN)$ is the *a priori* probability that a sample of the picture is a skin sample. It depends on the type of the studied sequence, e.g. it is greater for a videoconference sequence than for a landscape sequence. $P(\overline{SKIN}) = 1 - P(SKIN)$ as $SKIN$ and $\overline{SKIN}$ are complementary events. The conditional density probability function $T(C_b, C_r \mid SKIN)$ and $T(C_b, C_r \mid \overline{SKIN})$ have been estimated from (C_b, C_r) histogram of a large number of $SKIN$ or $\overline{SKIN}$ training samples. Assumption has been made that $T(C_b, C_r \mid \overline{SKIN})$ is a gaussian density probability function, product of two independent density probability function (one for C_b, the other for C_r). For $T(C_b, C_r \mid SKIN)$, it has been necessary to decorrelate C_b and C_r datas. Then, the decorrelated datas (Y_1, Y_2) have been assumed to be independent in order to estimate the joint density probability function by the product of two functions, $T(Y_1)$ and $T(Y_2)$. Tests have resulted in the following density probability function:

$$T(C_b, C_r \mid \overline{SKIN}) = \frac{exp(-(C_b-121.6)^2)/1136)}{\sqrt{1136 \cdot \pi}} \cdot \frac{exp(-(C_r-133.6)^2)/1132)}{\sqrt{1132 \cdot \pi}}$$

$$T(C_b, C_r \mid SKIN) = T(Y_1, Y_2 \mid SKIN) = T(Y_1 \mid SKIN) \cdot T(Y_2 \mid SKIN)$$

$$\begin{aligned} &with \\ &Y_1 = -0.595 \cdot (C_b - 99) + 0.804 \cdot (C_r - 160) \\ &Y_2 = -0.804 \cdot (C_b - 99) - 0.595 \cdot (C_r - 160) \\ &and \\ &T(Y_1 \mid SKIN) = \begin{cases} \frac{exp(-(Y_1+8)^2)/98)}{11.5 \cdot \sqrt{2 \cdot \pi}} \; if \; Y_1 \leq -8 \\ \frac{exp(-(Y_1+8)^2)/512)}{11.5 \cdot \sqrt{2 \cdot \pi}} \; if \; Y_1 \geq -8 \end{cases} \\ &T(Y_2 \mid SKIN) = \frac{exp(-Y_2^2)/80)}{\sqrt{80 \cdot \pi}} \end{aligned} \tag{3}$$

The rule expressing the criterion is:

- If skin probability is HIGH 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 skin probability is LOW then interest is (SOMEWHAT) LOW.

For the **motion criterion**, there are three input variables. The first one is a measure of the motion of the studied area. The corresponding 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. This is similar to the use of a non-singleton fuzzifier.[8]

The second variable is the mean value of the picture gradient. This gradient is expressed in each pixel and is the mean of the absolute difference of the pixel with its neighbours. The linguistic variable is called "GRADIENT". This value does not take into account the gradient distribution over the region and does not make any difference between an area covered by high spatial frequency texture and a region containing a small, high-contrasted object. Yet, it is highly probable that, while displaying the scene, the rendering of the small, high contrasted object is more important than the rendering of a random texture for the semantic understanding of the scene.

So, a third variable, named "ACTIVITY", denotes this notion. It is a value resulting from a morphological treatment of the gradient picture. Actually, the activity picture is deduced from the comparison of the closing of the gradient picture and of the opening of this closing. The activity value is equal to the gradient

value in the position where the difference between the difference between the closing and the opening of the closing is above a chosen treshold (2 has been chosen). Activity has high values for isolated high gradient values, corresponding in general to the border between contrasted areas.

Having defined the input variables, five rules express the motion and contrast criterion:

- If motion is HIGH then interest is HIGH.
- If motion is SMALL and activity is SMALL then interest is LOW.
- If activity is VERY HIGH then interest is SOMEWHAT HIGH.
- If motion is MEDIUM and activity is HIGH then interest is VERY HIGH.
- If motion is (VERY) HIGH and gradient is (SOMEWHAT) HIGH then interest is LOW.

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 follows:

- 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. In the first tests presented hereafter, the tracking is based on the improbable hypothesis that there is 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 pixel by pixel difference between the two pictures.

B) Fuzzy Sets Definition

Despite the fact that the rules can be clearly formulated, their interpretation still remains largely subjective. This subjectivity will be formalized through the following fuzzy sets definition. Fuzzy membership functions can have different shapes depending on the preference or experience of the designer. In practice, however, the most usual fuzzy systems shapes are the triangular and the trapezoidal ones. They simplify the required computation and are sufficient to help capture the modeler sense of fuzzy number. Such shapes are thus used in the present formalization. Let us now enumerate the involved variables and the attached fuzzy sets definitions:

(a) **"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).

(b) **"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).

(c) **"FRACTION BOUNDARY 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)).

(d) **"SKIN PROBABILITY"** Its universe of discourse is also [0,1]. The terms are LOW and HIGH. Figure 1 (d) illustrates them.

(e) **"GRADIENT"** The mean value used as the input value is divided by a value supposed to be higher than nearly all gradient values (30 has been used). Then it is truncated to reduce its universe of discourse to [0,1]. The terms HIGH and LOW are defined by the triangular shapes of figure 1 (e.1).

(f) **"ACTIVITY"** The mean value used as the input value is divided by a value supposed to be higher than nearly all activity values (30 has been used). Then it is truncated to reduce its universe of discourse to [0,1]. The terms HIGH and LOW are defined by the triangular shapes of figure 1 (e.2).

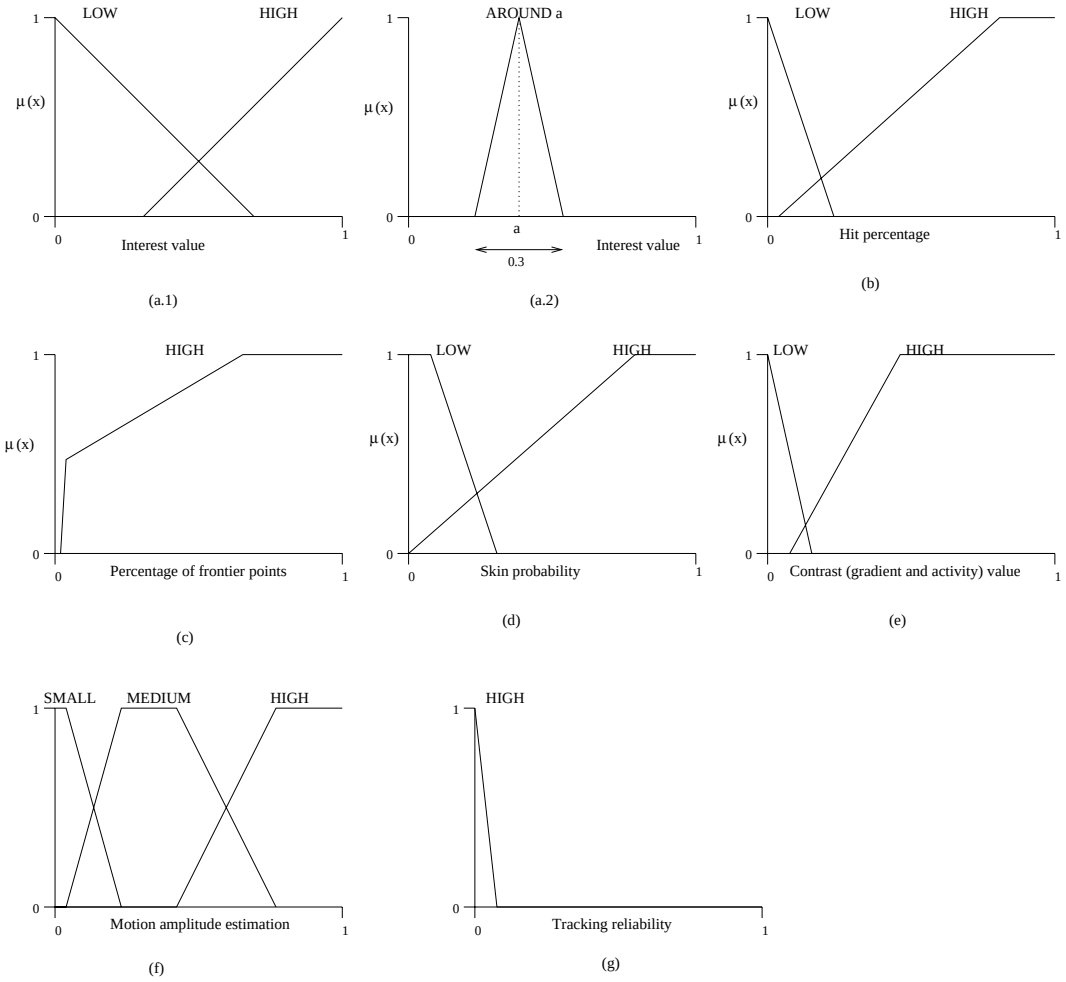

Figure 1: Representation of the membership functions defining fuzzy sets for each used linguistic variable.

(g) **"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).

(h) **"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).

The hedges used to transform the surfaces of these membership function are VERY or SOMEWHAT. VERY corresponds to the *square* operator while SOMEWHAT is defined by the *square root* operator.

2.1.3 FLS Implementation

In the first implementation, *"max"* and *"min"* have been chosen as the *t-norm* and *t-conorm* operators while the *minimum implication rule* was chosen for fuzzy reasoning.[7] The chosen fuzzifier transforms a crisp number into a fuzzy number whose membership function is a triangular function centered upon the crisp value. In the

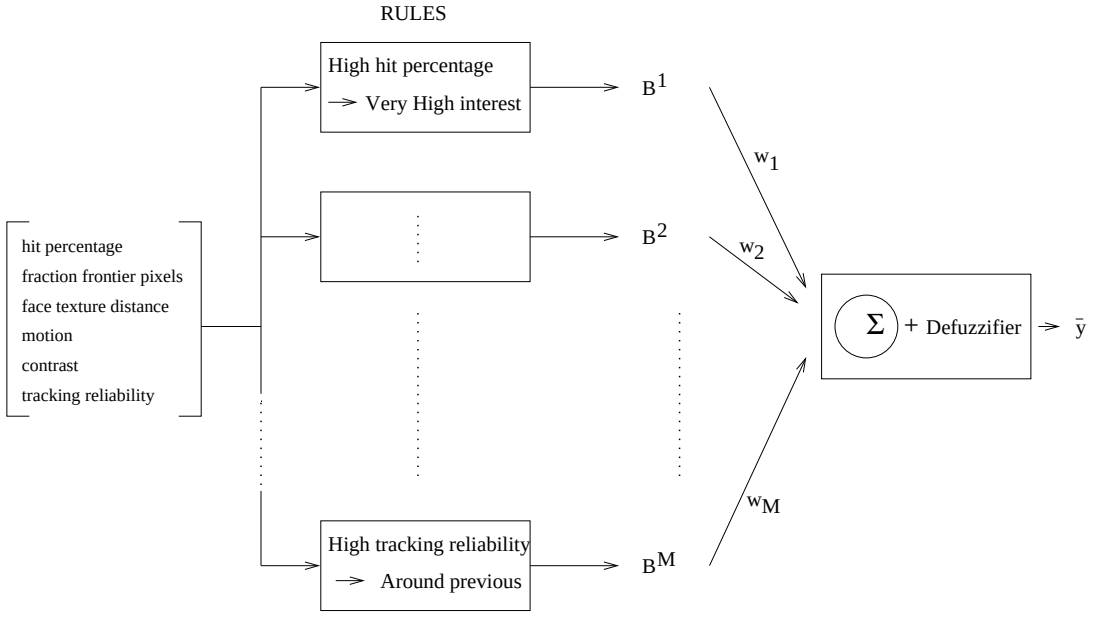

Figure 2: Fuzzy Logic System Architecture. The rules work in parallel. As the used defuzzifier is the height defuzzifier, there is no real distinction between the summation stage and the defuzzification stage. Both actions act at the same stage.

application, the fuzzifier has been used to fuzzify the motion estimated value. It would of course be more efficient to bind the triangle width to the uncertainty of motion estimation but this would require an accurate estimator of the motion reliability. The noise perturbing the motion estimation is just taken into account by a constant spreading of the triangular function. For the aggregation process, it seems obvious that the *additive aggregation* technique will be more effective than the *maximum combination* technique.[7,8] Indeed, the maximum combination does not allow an effective cooperation of the rules because rules with a low output would have no incidence on the result. Moreover it can be demonstrated[13] that, as the number of combined fuzzy sets increases, this combination technique tends to produce a uniform distribution for the resulting fuzzy set. For computation simplification purpose, the *height defuzzifier*[7] was chosen to combine the defuzzified values of each rule output subset. There was thus no need to perform a real aggregation of the output fuzzy sets, each set being defuzzified independently from the others. Moreover, such a defuzzification procedure permits to easily take into account the confidence 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*.[8] In case of several maximums, the lefter, righter or median one is chosen if the fuzzy membership function is respectively monotonic increasing, monotonic decreasing or non-monotonic.

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

$$\overline{y} = \frac{\sum_{i=1}^{M} w_i \mu_{B^i}(\overline{y}^i)\overline{y}^i}{\sum_{i=1}^{M} w_i \mu_{B^i}(\overline{y}^i)} \tag{4}$$

The resulting system architecture is depicted on figure 2. The example of figure 3 illustrates how a peculiar rule is treated.

3 RESULTS

Before studying their interest, extraction of picture regions is requested. As told in section 2, the segmentation should correspond to an image-based picture understanding method and should aim at dividing the picture in regions corresponding to parts of real objects in the scene. For the following analysis purposes, the extracted areas should satisfy two main constraints. First, they should be as large as possible because the larger the amount of data taken into account for making the decision, the more the analysis will be efficient. Secondly, data used for a region have to be related to a single real object. Regions have thus to follow the real objects edges and must not be shared by two distinct real objects. Practically, for the results presented hereunder, regions of uniform spatial features are extracted. This is achieved in two steps:

- First, a watershed algorithm based on immersion simulations[14] divides the image in small regions whose edges include real objects edges. Oversegmentation inherent to this approach (see figure 4: Oversegmentation) can be reduced by tresholding the gradient used for the immersion process and by merging, during the flooding step, the neighbouring regions having areas lower than a particular treshold. Nevertheless, obtained results show that these tresholds have to be kept at a rather low value if one wants to avoid merging parts of distinct real objects in a unique region. Another way to reduce oversegmentation is to make use of markers.[15,16] This approach, usefull to generate large regions, remains very sensitive to the markers choice. A bad choice can lead to the loss of some edges between real objects. As the loss of some significant edges is worst than oversegmentation for the analysis purpose, markers will not be used.
- The second step tries to reduce the unavoidable oversegmentation due to the first step. It consits in merging adjacent regions having near spatial features. In practice, regions having near luminance and chrominance mean values are merged. This treatment permits to generate larger regions from the result of the first oversegmentation (figure 4: Segmentation after merging).

Figure 4 presents the results obtained for some MPEG4 sequences. In these pictures, darkness corresponds to interest. For each criterion, a figure illustrates the defuzzification output obtained from the aggregation of the rules of the studied criterion. It is noticeable that the combination of several criteria permits to strongly reduce erroneous interpretations. As an example, the "texture face criterion" is mislead by the color of Akiyo's jacket. Nevertheless, in the global result, other criteria are strong enough to overcome this local error.

4 CONCLUSION AND PERSPECTIVES

This paper has presented a way to allocate a "level of interest" to regions of coherent features extracted in a video sequence. In a first step, the picture has to be segmented in a way that should match as much as possible the "real" objects as detected by a human viewer. Then a procedure to quantify the subjective interest of the different image areas has been introduced through the implementation of a Fuzzy Logic System. Fuzzy logic has been chosen because of its ability to deal with the human subjectivity and to combine different cooperative opinions. Up to now, five criteria reflecting different pieces of subjective knowledge about "What is an area of interest?" are used.

First results are very promising. This tool is really able to distinguish interesting areas from less important ones in a wide range of video sequences. This result could be useful for very low bitrate video coding for example. Indeed, to achieve VLBR video coding, significant compression is required. In this context, an interesting approach is to extract the subjectively important areas of a scene and to allocate more bits to these areas than to other parts of the scene. This allows reduction of the amount of information to be transmitted, while keeping good subjective quality of images. Such a content-based bitrate allocation is now facilitated by to the new scope of

MPEG-4,[17] that includes a representation based on Video Object Planes (VOP) in its Verification Model.[18] In this context, the tool could permit to control an adaptative bit allocation by coding more coarsely the region that are of rather low interest.[19] Such a tool can also allow a major step in the direction of image comprehension; this is important because a semantic understanding of pictures could be useful for applications involving the extraction and manipulation of objects in a scene.

Nevertheless, significant improvements are still foreseenable. One improvement could be the refinement and completion of the criteria. As explained earlier, the motion and continuity criteria could be improved by an evaluation of the motion estimation reliability.

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 about the interest values to attach to the regions of the picture.

Although we have chosen the FLS rules according to the subjective knowledge, it could be interesting to identify FLS rules in a more objective way. Clustering methods for instance allow such identification from pairs of $(input, output)$ vectors.[20] The definition of such pairs 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[21] and would probably be an interesting way to objectively match a level of interest to the training data.

5 REFERENCES

[1] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing: Analysis and Machine Vision*. Chapman & Hall Computing, London, 1 edition, 1983.

[2] Toshiro Terano, Kiyoji Asai, and Michio Sugeno, editors. *Applications in Industry*, chapter 3, pages 140–161. ACADEMIC PRESS, 1994. ISBN 0-12-685242-1.

[3] J.C. Bezdek. Partition Structures: A Tutorial. In J.C. Bezdek, editor, *Analysis of Fuzzy Information*, volume 3, pages 81–107. CDC Press, Boca Raton, Florida, 1987.

[4] M.M. Trivedi. Analysis of Aerial Images Using Fuzzy Clustering. In J.C. Bezdek, editor, *Analysis of Fuzzy Information*, volume 3, pages 133–151. CDC Press, Boca Raton, Florida, 1987.

[5] K. Hirota, Y. Arai, and F. Hachisu. Robot for Recognition of Moving Objects and Replacement of Moving Objects Using Fuzzy Logic. pages 15–22. 2nd Fuzzy Systems Symposium of the Japan Chanter of IFSA, 1985. (in Japanese).

[6] T. Terano, S. Masui, S. Kono, and K. Yamamoto. Recognition of Crops by Fuzzy Logic. pages 374–377, Tokyo, July 1987. 2nd IFSA Congress.

[7] Jerry M. Mendel. Fuzzy Logic Systems for Engineering: A Tutorial. *Proceedings of IEEE*, 83(3):345–377, March 1995.

[8] 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, AP PROFESSIONAL edition, 1994. ISBN 0-12-194270-8.

[9] R.R. Yager and D.P. Filev. Template-based fuzzy systems modeling. *J. Intell. and Fuzzy Syst.*, 2:39–54, 1994.

[10] L.A. ZADEH. Fuzzy Sets. *Information and Control*, 8:338–353, 1965.

[11] M.P. Queluz and B. Macq. Signal-adapted motion compensation for video compression. Paris, September 1992. ISSES, URSI.

[12] W.K. Pratt. Photometry and Colorimetry. In *Digital Image Processing*, chapter 3, pages 72–79. A Wiley Interscience Publication, University of Southern California, 1978. ISBN 0-471-01888-0.

[13] Bart Kosko. The Dynamical System Approach to Machine Intelligence: Fuzzy Systems as Parallel Associators. In *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, chapter 1, pages 29–32. Prentice-Hall International, Inc., University of Southern California, 1992. ISBN 0-13-612334-1.

[14] L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, June 1991.

[15] F. Meyer and S. Beucher. Morphological segmentation. *Journal of Visual Communications and Image Representation*, 1(1):21–46, September 1990.

[16] P. Salembier. Morphological Multiscale Segmentation for Image Coding. *Signal Processing*, 38(3):359–386, August 1994.

[17] F. Pereira. MPEG4: A New Challenge for the Representation of Audio-Visual Information. pages 7–16, Melbourne, March 1996. Picture Coding Symposium.

[18] MPEG Video group. MPEG-4 Video Verification Model, version 3.0, ISO/IEC SC29WG11 (N1277). Tampere, July 1996.

[19] C. De Vleeschouwer, T. Delmot, X. Marichal, and B. Macq. A Fuzzy Logic System for Content-Based Bitrate Allocation. *Signal Processing: Image Communication*. Accepted for publication in the MPEG-4 special issue.

[20] 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.

[21] 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 - Journées d'études et d'échanges: Compression et représentation des signaux audiovisuels.

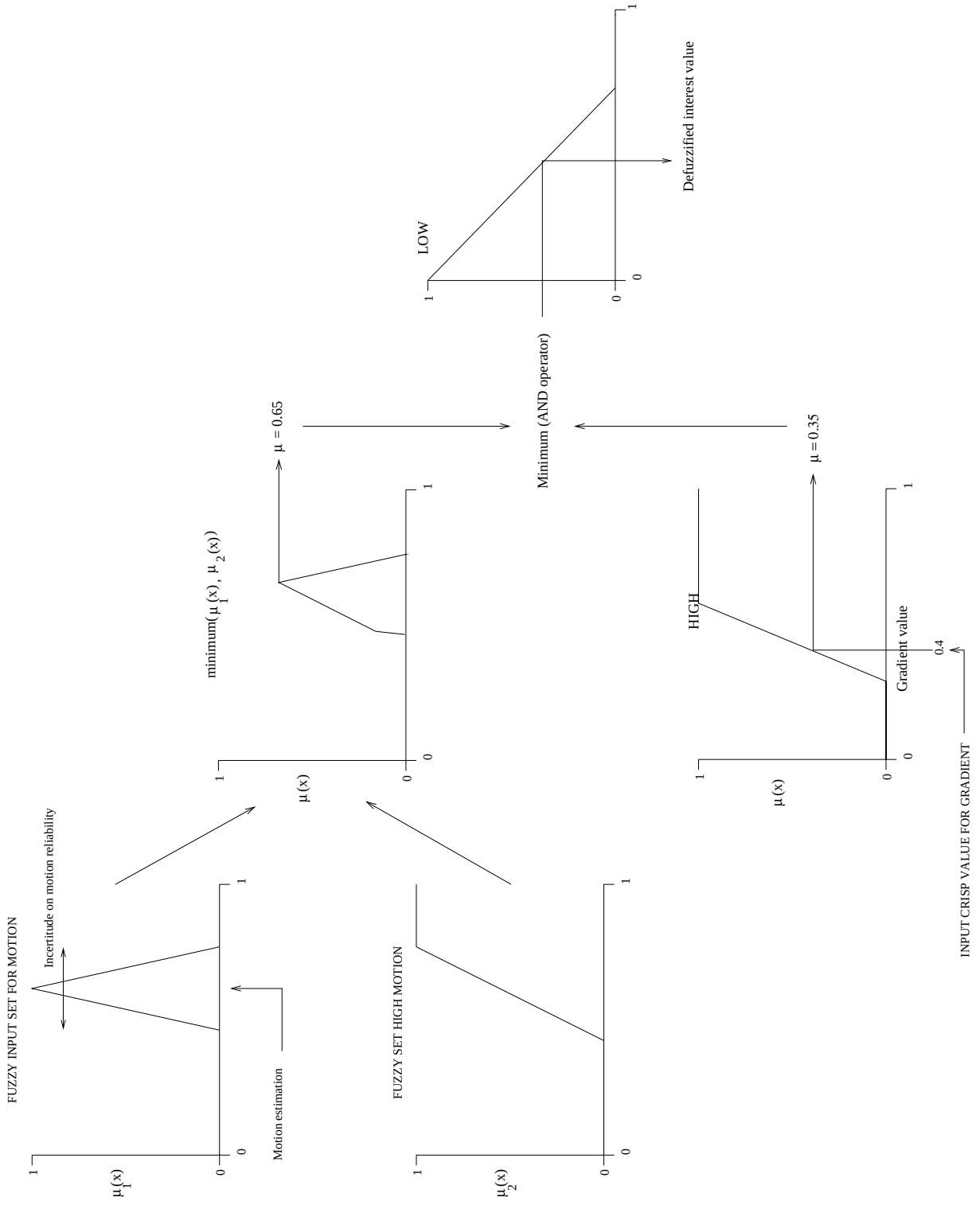

Figure 3: Example of the way rules are working. Motion rule: "If motion is HIGH and gradient is HIGH then interest is LOW".

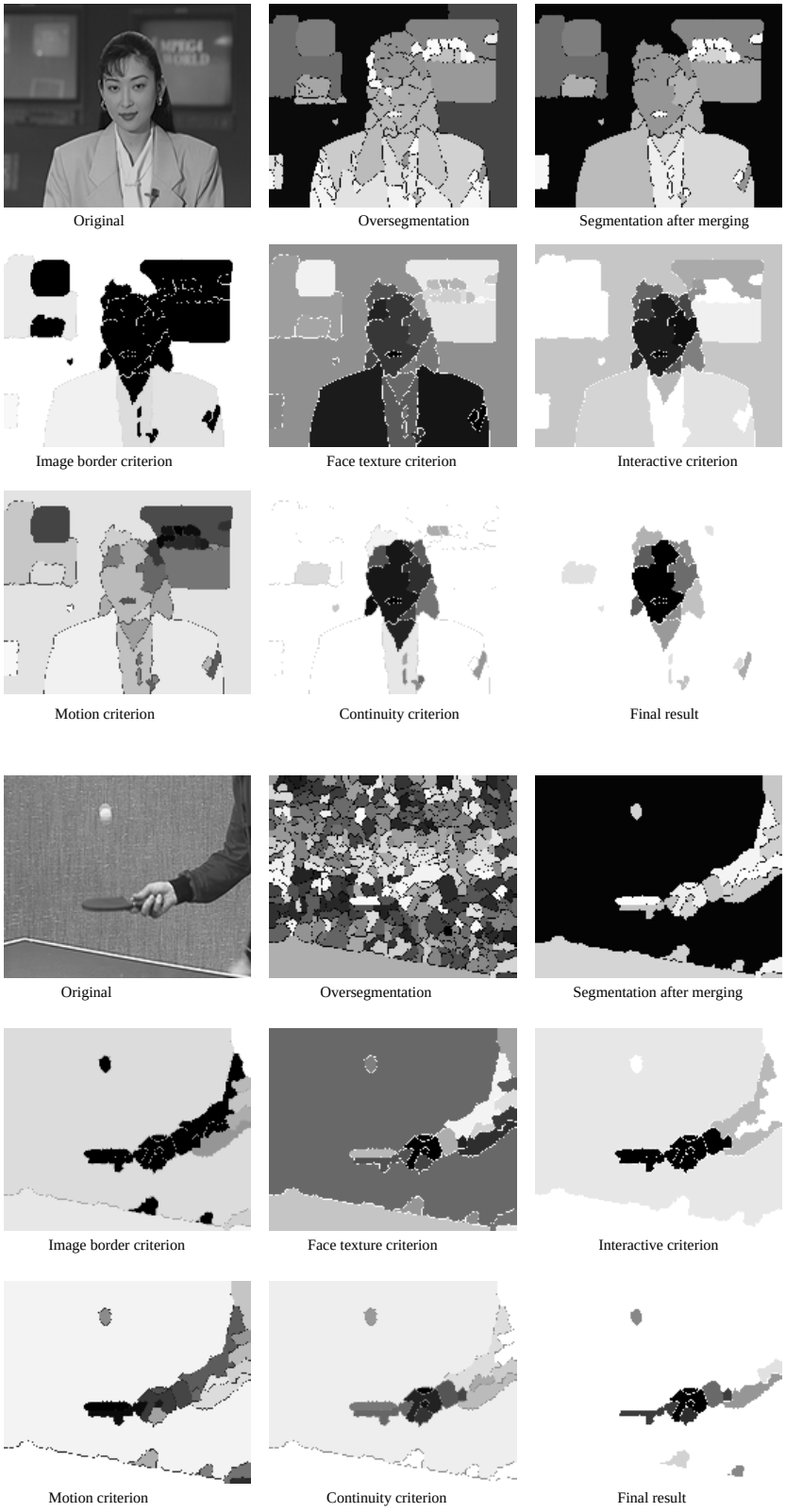

Figure 4: Results of a first implementation. Sequences "Akiyo" and "Table Tennis". The darkest regions are the most interesting ones. The results obtained for each criterion illustrate the defuzzification output obtained from the aggregation of the rules of the studied criterion.