ELSEVIER

Signal Processing: *Image Communication* 10 (1997) 115–141

SIGNAL PROCESSING: IMAGE COMMUNICATION

A fuzzy logic system for content-based bit-rate allocation

C. De Vleeschouwer [*,1], T. Delmot [2], X. Marichal [3], B. Macq

Laboratoire de Télécommunications et Télédétection, Université Catholique de Louvain, Bâtiment Stévin – 2, place du Levant, B-1348 Louvain-la-Neuve, Belgium

Abstract

When video coders communicate at very-low bit-rate, it is often difficult for them to preserve an acceptable global quality of the images. A selection of the key regions according to the semantic relevance may therefore be useful for improving the quality, as it would permit an adaptive bit allocation, which is important for good subjective quality at (very-)low bit-rate: the essential features could be extracted and coded with a good quality, while the remaining portions of the image would be coarsely transmitted. This paper describes tools to perform classification of regions according to the subjective priority, and a generic algorithm to optimally share out the bit-rate. Among others, an important contribution of this paper is the use of *Fuzzy Logic* in order to model human subjective knowledge about the attribution of priorities to picture regions. An original bit-rate allocation method is also proposed as a solution to the optimization problem posed by the rate-distortion theory. The paper concludes presenting some preliminary results. © 1997 Elsevier Science B.V.

Keywords: Bit-rate optimization; Low bit-rate; Video analysis; Content-based scalability; MPEG-4 functionalities; Rate-distortion theory; Video object plane (VOP); Fuzzy logic

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.

For picture coding purposes, analysis tools have to be as generic as possible because of the large variety of images they have to deal with. Up to now, block transform coding (JPEG [31], MPEG [9, 19]) use a frequency analysis of local properties while object-based algorithms ([18], MORPHECO [15], SESAME [4]) exploit a texture and contour analysis. These analysis procedures mainly aim at decorrelating the picture information. They also take more or less into account the visual acuity in order to improve the coding

* Corresponding author. Tel.: +32 10474105; fax: +32 1047 2089; e-mail: devlees@tele.ucl.ac.be.

[1] Research of C. De Vleeschouwer is supported by a grant F.N.R.S. – Alcatel Bell.

[2] Research of T. Delmot is supported by Belgacom.

[3] Research of X. Marichal is supported by a grant of the FRIA.

PII S0923-5965(97)00021-0

efficiency. Unfortunately, the achieved compression rates may not be sufficient for very-low bit-rate (VLBR) transmission of image sequences and degradations of the picture may subsequently result.

To achieve VLBR video coding, significant compression is required. In this context, an interesting approach could be 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 would allow reduction of the amount of information to be transmitted, while keeping good subjective quality of images. Such a content-based bit-rate allocation is now facilitated by to the new scope of MPEG-4 [19], that includes a representation based on Video Object Planes (VOP) in its Verification Model [17].

The goal of this paper is to introduce an algorithm for *content-based bit-rate allocation* in the framework of VLBR video coding. An automatic tool able to analyze the different areas and to attribute them a 'level of interest' is first introduced in Section 2. This tool is effective for a very large range of natural pictures. It is versatile and flexible in that it can be used standalone or can take into account an a priori knowledge about the scene. 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. Conceptually, there are two major approaches to image understanding [24]:

- *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 [26]: image recognition using c-means clustering [2,28], recognition of moving objects [6] and recognition of vegetables using linguistic expressions [27] 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.

We then discuss the problem of bit-rate allocation as it appears that no general analytic equation is available to address this arduous problem. With the help of *rate-distortion theory*, we propose a practical algorithmic solution. Two assumptions are made: both a classification of the regions according to their semantic importance and a quality criterion exist. The algorithm (Section 3) makes use of a buffer in order to optimally distribute the bits with direct emphasis on the most important parts of the picture. Of course, if the bit-rate is sufficient, it tries to give all the regions a good quality.

Section 4 presents some results. Exhaustive results are provided for the region classification. It is demonstrated that the tool performs correctly even with over segmentation inherent to current segmentation tools. A case study of the allocation scheme is also introduced. Finally, preliminary results of coding with the MPEG-4 VM [17] are exposed. In this cases, the segmentation has been hand guided in order to obtain coherent VOPs.

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 4.1 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.

In order to perform 'subjective video coding', a tool able to classify the regions of a picture according to their relevance is needed. The quality requirements being directly related to the semantic interest, the meaningful features of a scene must be preserved through the coding process. A way to reach this aim is to maintain high quality coding in areas whose characteristics are important while allowing lower quality in regions whose characteristics are less important.

A criterion of interest is thus searched for. It must be able to automatically attribute a level of interest (related to the required quality level) to the various parts of a picture. For the purpose of dealing with semantic interpretation of images, a single criterion is clearly insufficient. Obviously, that 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, [12, 5]). 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 [12] that an 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 [32], implications rules [33], the aggregation method, the fuzzifier and the defuzzifier [5].

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. It is obvious that the human brain, thanks to its a priori knowledge of real-world 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 temporal components specific to moving images.

2.1.1.1. 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.

2.1.1.2. Inter-images criteria

Motion criterion: It is based on a succession of motion estimations (using [21]) between the images at the highest frequency (even if the sequence is coded at 8 Hz, the criterion will compute three motion estimations between images at 25 Hz), it reduces the interest of regions with no motion and a very small gradient (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).

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

2.1.2.1. 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 referred as '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 is the geodesic distance [30] between a kernel of points corresponding to face texture samples and the point representing the region in a chrominance space. A desirable property for the used color coodinate system could be that any unit change in the chromaticity diagram of that system should be perceived by an observer as an equivalently noticeable color shift. The CIE *Uniform Chromaticity Scale Color Coordinate System* (= (L,u,v) system) is an attempt for obtaining such a system [20]. 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. Fig. 1 illustrates the concept of geodesic distance by a map representing, in the chrominance space, a set of curves equidistant from the kernel. The handled input

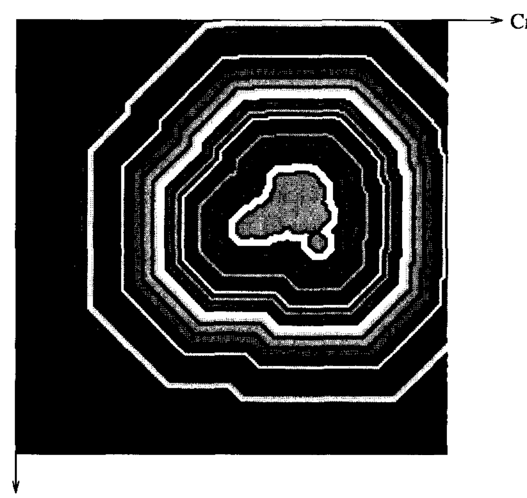


Fig. 1. Representation, in the chrominance space, of curves equidistant from the face texture samples kernel.

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 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 [5].

The second variable is a measure of the area contrast. The linguistic variable 'CONTRAST' denotes this notion. Up to now, the used value has been a mean value of the picture gradient.

Remark. This choice does not seem to be the most relevant as it does not take into account the contrast 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.

Having defined the input variables, 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.
- 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 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.

2.1.2.2. 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 in Figs. 2(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 Fig. 2(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 (Fig. 2(c)).
(d) FACE TEXTURE DISTANCE: The distance used has been normalized by its maximal value.

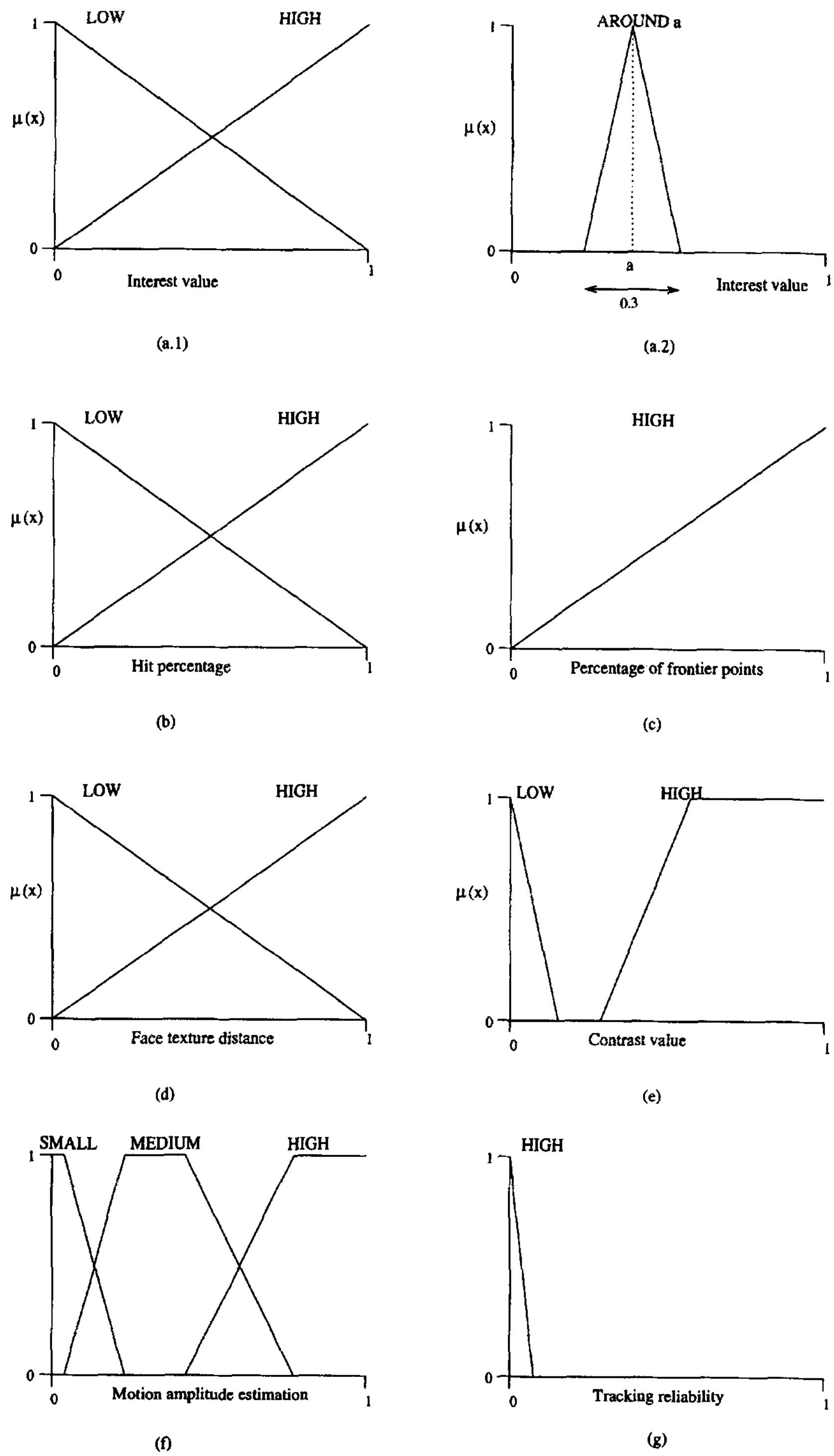


Fig. 2. Representation of the membership functions defining fuzzy sets for each used linguistic variable.

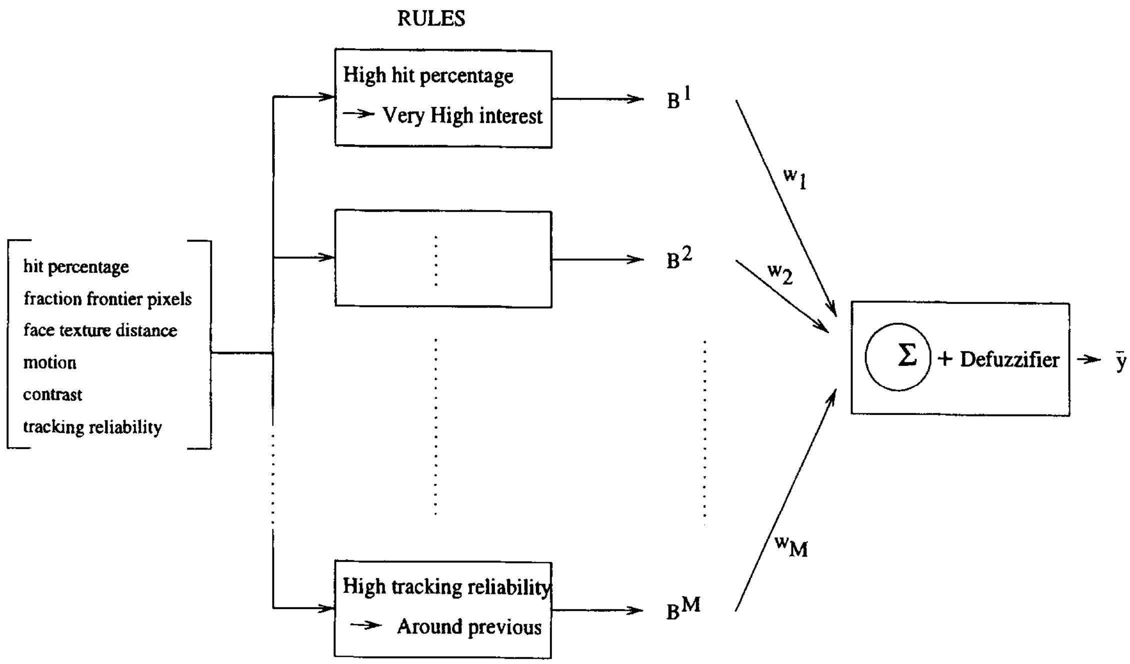


Fig. 3. Fuzzy Logic System Architecture. The rules work in parallel. As the used defuzzifier is the height defuzzifier, there is no real distinction between the summation stage and the defuzzification stage. Both actions act at the same stage.

So, its universe of discourse is also [0,1]. The terms are LOW and HIGH. Fig. 2(d) illustrates them.

(e) 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 Fig. 2(e).

(f) 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 in Fig. 2(f).

(g) 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 Fig. 2(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 [12].

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 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 [12, 5]. 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 [7] that, as the number of combined fuzzy sets increases, this combination

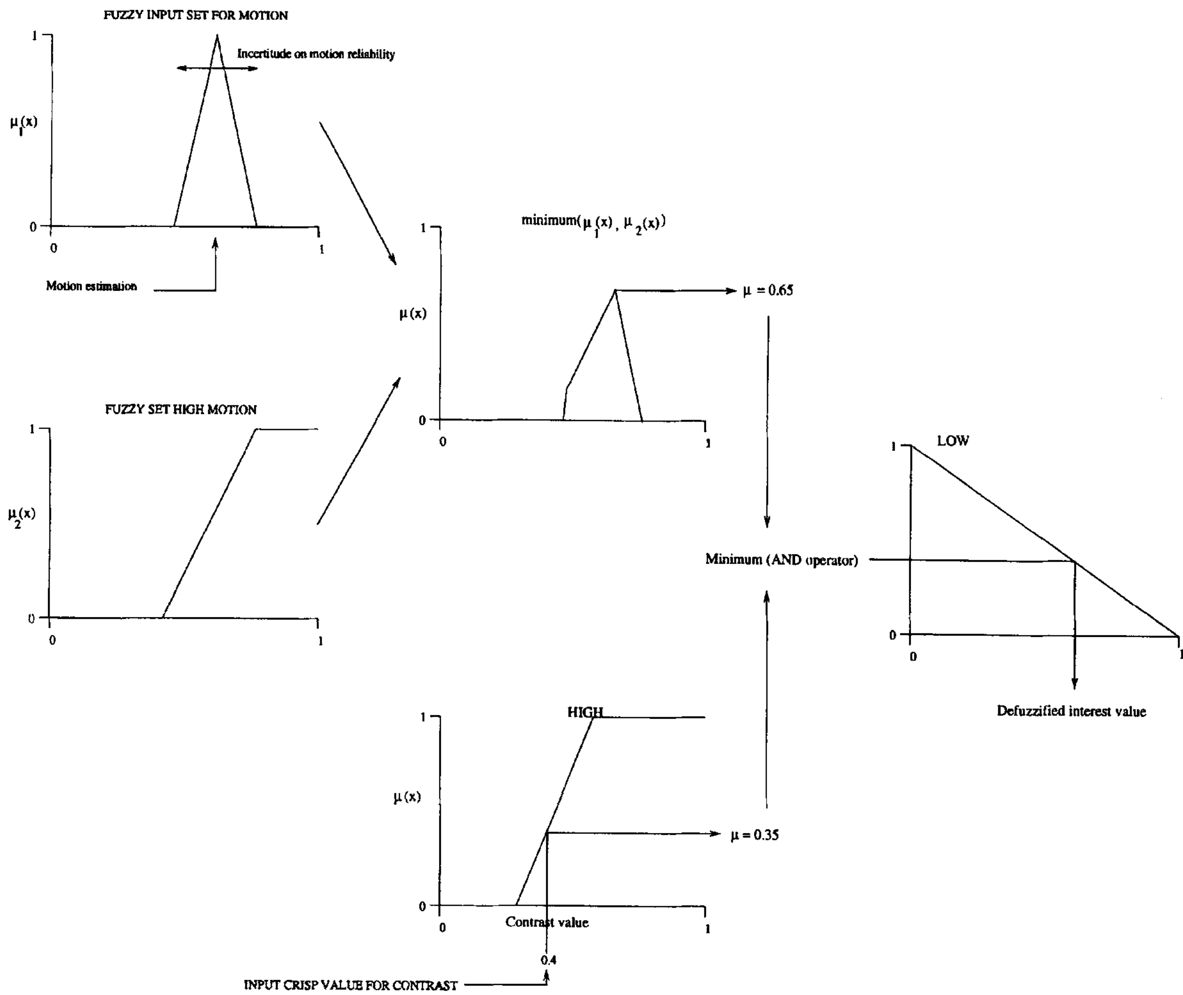


Fig. 4. Example of the way rules are working. Motion rule: "If motion is HIGH and contrast is HIGH then interest is LOW".

technique tends to produce a uniform distribution for the resulting fuzzy set.

For computation simplification purpose, the *height defuzzifier* [12] 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* [5]. In case of several maximums, the leftmost, rightmost 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

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

The resulting system architecture is depicted in Fig. 3. The example of Fig. 4 illustrates how a peculiar rule is treated.

3. Optimum bit-rate allocation

The bit-rate allocation issue in video coding is generally viewed as a basic single variable regulation scheme applied over entire images. For instance, the MPEG-2 [9] regulation acts on the DCT quantization step in order to regulate the amount of bits the coder generates for each image. This quantization step is used to code every part of the image and only the global quality of the images can be changed over the video sequence. This bit-rate control method has proved to be efficient for medium to high bit-rate coding, where the global quality can be maintained at a good level. The transmission bandwidth allows to spend as many bits as required. However, for (very-) low bit-rate (VLBR) coding and especially the bit-rates allowed for wireless communications (i.e. 5 to 15 kbits/s), such global and mono-parameter bit-rate regulation schemes reach their limit: it appears very expensive to maintain a constant quality on all parts of the image whether or not the part is interesting for the user. The bandwidth is so narrow that the global image quality becomes very poor and annoying artifacts arise.

In this section, we describe a new way to perform the bit-rate regulation within the framework of VLBR video coding. We assume that not all images parts need to be displayed with the same quality, and introduce this hypothesis in the framework of the *Rate-Distortion theory*. When the transmission bandwidth is too narrow to ensure a good global quality over the whole image, it becomes relevant to select crucial areas of the image where good quality will be maintained while other areas are coarsely encoded.

We first start by setting the adequate hypotheses in order to reach an analytic expression for regulating the content-based quality. This is made through the rate-distortion theory. Since an exact analytic solution seems difficult, an algorithm that approximates the optimum is proposed.

3.1. Region classification and image partitioning

3.1.1. Region classification: a matter of subjective priority

If one considers $I(x, y, t)$, an original image at time t, a segmentation algorithm provides a segmentation R of I in N regions:

$$R(I(x,y,t)) = \{R_1, R_2, \ldots, R_N\},$$
$$\bigcup_{i=1}^{N} R_i = \text{image}\, I, \qquad R_i \bigcap R_{j_{(j \neq i)}} = \emptyset. \tag{2}$$

The regions R_i are distinct (they do not overlap) and they cover the whole image. As mentioned in the preliminary remark of Section 2, the better the segmentation R matches the human image understanding in terms of pictured objects, the better it will fit the actual goal of quality modulation according to the subjective importance the end-user will attribute to the different regions.

Thanks to the region classification (Section 2), every region R_i is attributed a label, namely a 'level of interest', l_{R_i} that tries to objectively describe how important is the region according to human subjective criteria. It may happen, without any loss of generality, that $l_{R_i} = l_{R_{j_{(j \neq i)}}}$. If R_i is subjectively more important than R_j then $l_{R_i} > l_{R_j}$.

3.1.2. Coding partition

At the same time, the coder makes use of a partition P of M elements that have properties similar to the ones of the segmentation:

$$P(I(x,y,t)) = \{P_1, P_2, \ldots P_M\},$$
$$\bigcup_{k=1}^{M} P_k = \text{image}\, I, \qquad P_k \bigcap P_{l_{(l \neq k)}} = \emptyset. \tag{3}$$

The partition is truly devoted to the coding process and each of its element represents the smallest entity that is possible to be encoded on its own. If one discards the overhead and multiplex information, one may say that every partition element P_k receives separately its own bits. If b_{P_k} depicts the number of bits assigned to the coding of P_k, there exists a minimal amount of bits, $b_{\min,P_k}$, requested to encode the element with a minimal quality.

In order to clarify the distinction between R and P, three coding schemes are briefly reviewed. In all three cases, the initial segmentation R is the same. However, one additional constraint must be added to Eq. (2): in order to allow a bit-rate partitioning (on the partition P) via the levels of interest (attached to the segmentation R), the segmentation must be altered so that the frontiers between regions R_i and R_j are placed on frontiers between partition elements.

3.1.2.1. H.263 partition. In the H.263 [25] based video coding, each partition is composed of a set of 16 × 16 pixels macroblocks. $b_{\min,P_k}$ is due to the need to at least transmit the DC component (for intra macroblocks) of the DCT transform, or the motion vector (for inter macroblocks) related to the block. It appears obvious that restricting the segmentation to blocks of 16 × 16 pixels has an impact on the flexibility and the precision of the adaptive bit-rate allocation over regions R_i.

3.1.2.2. Multiscale partition. When using a multiscale partition, like the one used in [10, 11], the gain comes from the fact that the block partition will adapt itself to the object contours. The selective distribution of the bit-rate will then be more efficient because the partition P will better match the segmentation R.

3.1.2.3. Relation with MPEG-4 VOP concept. Considering the last developments of the MPEG-4 Video Verification Model [17], the subjective classification designed on the regions R_i integrates itself very well. Within this Verification Model, Video Object Planes (VOP) are defined as two-dimensional video objects (i.e. a collection of pixels extracted from the original input image) that are coded separately and multiplexed into the final video bitstream. Using the previously described region ranking system, where R determines the VOPs shape and $I(x, y, t)$ their texture, it then appears possible to develop a scheme for VOP generation. The bit-rate regulation scheme can directly be applied on each VOP. For instance, the coding process could be initiated with a different quantization step for every VOP, this quantization step being determined by the level of interest l_{R_i}. Of course, in order to reduce the multiplex information, it may be useful to assemble in each VOP all the regions R_i having received an identical level of interest.

3.2. Content-based bit-rate allocation

In the following, the chosen coding algorithm will be represented by a letter T. In relation to this coding algorithm, a set of coding and quantizing parameters, say vector $\overline{p}$, has to be chosen. The amount of bits necessary to code a partition element P_k with the algorithm T will of course depend on the chosen parameters $\overline{p}_{P_k}$. So, from now on, b_{P_k} will be noted $b_{P_k,\overline{p}_{P_k}}$.

The problem now consists in attributing every partition element P_k some coding parameters $\overline{p}_k$, under the following constraints:
- Every partition requires a minimal amount of bits ($b_{P_k,\overline{p}_{P_k}} \geqslant b_{\min,P_k}$) in order to ensure the minimal quality.
- The image is allocated a maximum number of bits B: $\sum_{k=1}^{M} b_{P_k,\overline{p}_{P_k}} \leqslant B$. Of course, B may vary from one image to another according to the amount of bits already used for the previous images.
- The priority, in terms of resulting quality, must be given to partition elements whose pixels belong to a region R_i with a high level of interest l_{R_i}.

Let us consider a function f taking as arguments the image itself, the coding algorithm T, the segmentation along with the regions classification and the image partitioning whose variables are the different coding parameters which influence the bit allocation to every partition element. If, in addition, the function f returns a value describing how much the various constraints are met and the quality is high, the allocation problem consists in finding the parameters that maximize this function:

$$\begin{bmatrix} \overline{p}_{\mathrm{opt},P_1} \\ \overline{p}_{\mathrm{opt},P_2} \\ \vdots \\ \overline{p}_{\mathrm{opt},P_M} \end{bmatrix} = \underbrace{\arg\max}_{\overline{p}_{P_k}} f\left(I(x,y,t), T, \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_M \end{bmatrix}, \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{bmatrix}, \begin{bmatrix} l_{R_1} \\ l_{R_2} \\ \vdots \\ l_{R_N} \end{bmatrix}, B, \begin{bmatrix} \overline{p}_{P_1} \\ \overline{p}_{P_2} \\ \vdots \\ \overline{p}_{P_M} \end{bmatrix} \right) \tag{4}$$

under the constraints

$$\sum_{k=1}^{M} b_{P_k,\overline{p}_{P_k}} \leqslant B \quad \text{and} \quad \forall k\colon\ b_{P_k,\overline{p}_{P_k}} \geqslant b_{\min,P_k}. \tag{5}$$

3.2.1. Quality evaluation

The analytic function f must verify that the priorities in terms of interest between the different regions are met. It must also ensure that the chosen bit-rate provides the best possible quality. To achieve these two goals, a way of evaluating the subjective quality of (parts of) images is needed. To be in agreement with the levels of interest, the quality criterion used must also be based on the quality perception model of the Human Visual System (HVS). Improvements have already been reached in that field, and solutions are proposed in [3, 29].

In the following, it is considered that a quality function Q evaluating the subjective quality of a coded picture $\hat{I}(x,y,t)$ with regards to the original image $I(x,y,t)$, over the region R_i exists:

$$Q(I(x,y,t),\hat{I}(x,y,t),R_i). \tag{6}$$

Obviously, $Q(I(x,y,t),\hat{I}(x,y,t),R_i)$ will depend on the parameters $\overline{p}_{P_k}$ used to encode the partition elements included in region R_i.

3.2.2. Bit-rate allocation in the framework of rate-distortion theory

A classical framework for bit allocation is the Rate-Distortion (RD) theory [1, 22], which deals with the minimization of source distortion D for a given coding bit budget. Such a theory may help better formalizing Eq. (4). Classically, RD uses a quality function Q (see Section 3.2.1) in order to evaluate the distortion D and the problem becomes:

> Given an image $I(x,y,t)$, a coding bit-rate B and a quality function Q to evaluate the distortion, the constrained problem consist in attributing every partition element P_k of I some coding parameters $\overline{p}_k$ such that
>
> $$Q=\sum_{k=1}^{M} Q(I(x,y,t),\hat{I}(x,y,t),P_k) \tag{7}$$
>
> is maximum under the constraints formulated by (5).

This problem has already been treated exhaustively in the literature. In [29], such rate-distortion equation is used to provide joint optimal segmentation and coding. Our intention is to provide a bit-rate allocation method able to take into account the requirements in term of priorities resulting from the FLS classification tool.

3.2.2.1. One goal optimization. A way to realize the allocation in this framework is to solve the following optimization problem:

> Given an image $I(x,y,t)$, a bit-rate B, a coding partition P of M elements P_k, a segmentation R of N regions R_i, and a 'level of interest', l_{R_i} that tries to objectively describe how important is the region R_i according to human subjective criteria, the constrained problem consists in attributing every partition element P_k some coding parameters $\overline{p}_k$ such that
>
> $$Q=\sum_{i=1}^{N} l_{R_i} Q(I(x,y,t),\hat{I}(x,y,t),R_i) \tag{8}$$
>
> is maximum under the bit-rate constraints of Eq. (5).

Thanks to the presence of the factors l_{R_i}, the optimum will favor regions of high priority level for the obtention of high quality. In order to verify this assertion, the constrained problem is converted into an unconstrained problem by means of a Lagrange multiplier $\lambda \geqslant 0$. The Lagrangian cost to maximize becomes

$$J(\lambda)=\sum_{i=1}^{N} l_{R_i} q_{R_i} + \lambda\left(B-\sum_{i=1}^{N} b_{R_i}\right), \tag{9}$$

with $b_{R_i} = \sum_{k|P_k\in R_i} b_{P_k,\overline{p}_{P_k}}$ and $q_{R_i} = Q(I(x,y,t), \hat{I}(x,y,t),R_i)$.

By extension of the solution to a continuous problem, at the equilibrium, the solution to the discrete problem provides q_{R_i} such that

$$l_{R_i}\frac{\Delta q_{R_i}}{\Delta b_i} \approx \lambda^* = \text{constant value.} \tag{10}$$

The gradation of the qualities of the regions according to their level of interest appears thus strongly related to the behavior of the function $q_{R_i}(b_{R_i})$ that is of course increasing. Moreover, in the framework of video coding, further improvements of the quality always become more expensive as the quality increases. So, $\Delta q_{R_i}/\Delta b_{R_i}$ decreases when q_{R_i} increases. Thanks to this particularity, the maximization problem will provide a solution

- which is strongly advantageous to the regions of high interest when the available bit-rate (and thus the reached quality) is low,
- which tends to progressively uniformize the qualities of the different regions as the reached quality becomes higher.

This behavior corresponds to our purpose. Nevertheless, it suffers from a major drawback: the balance of quality between interesting and non-interesting regions as a function of the regions priorities will strongly depend on the coding algorithm T that fixes the function $q_{R_i}(b_{R_i})$. No real control exists on the incidence of the level of interest provided by the FLS on the quality gradation.

3.2.2.2. Two goals optimization. In order to better take into account the priority classification resulting from the FLS, one has to modify the rate-distortion function and to consider the allocation problem as a *double objective problem*. Under a budget constraint, two goals have to be reached:

- a maximum quality; and
- a quality distribution according to the level of interest of the various regions.

The second objective deserves some explanations. In order to respect the hierarchy of quality among the different regions, quality can be made proportional to the respective priority level, i.e. to ensure

$$\frac{q_{R_i}}{q_{R_j}} \approx \frac{l_{R_i}}{l_{R_j}} \quad \forall i,j.$$

The resulting optimum tends thus to aline each couple (q_{R_i}, l_{R_i}) on a straight line in the corresponding space (Fig. 5(a)). Of course, the line slope increases with the bit-rate.

However, it appears important to reach a good homogeneous quality all over the image when the bit-rate is large enough. This is made possible by introducing a parameter $\theta > 0$ as represented in Fig. 5(b). Actually, θ relaxes the origin of the straight line that has no more to start at the origin. The aim is to satisfy

$$\frac{q_{R_i} - \theta}{q_{R_j} - \theta} \approx \frac{l_{R_i}}{l_{R_j}} \quad \forall(i,j).$$

Considering θ as an increasing function of the quality of the region of highest priority (i.e. $q_{\max(l_R)} = q_{R_j} \mid l_{R_j} \geqslant l_{R_i}\ \forall i$), the optimum tends towards a straight line whose slope becomes lower as quality increases.

Taking these two objectives into account, the Lagrangian cost function to maximize becomes

$$J(\lambda) = \sum_{i=1}^{N} q_{R_i} - \beta \sum_{i,j} \left| \frac{q_{R_i} - \Theta(q_{\max(l_R)})}{q_{R_j} - \Theta(q_{\max(l_R)})} - \frac{l_{R_i}}{l_{R_j}} \right| + \lambda \left(B - \sum_{i=1}^{N} b_{R_i} \right), \tag{11}$$

where β is a weighting factor that manages the trade off between the two goals and $\Theta(q)$ is the increasing function that allows to obtain homogeneous quality at high bit-rates.

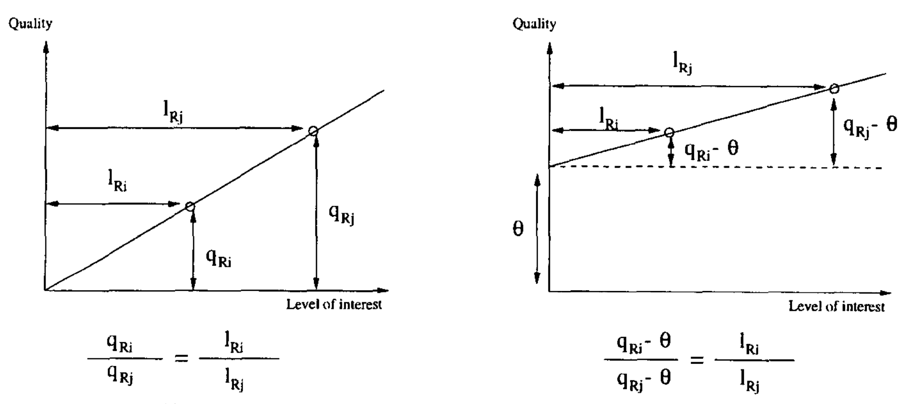


Fig. 5. Visualisation of the objective expressing the required quality gradation in terms of interest gradation: (a) quality proportional to interest, (b) generic linear relation.

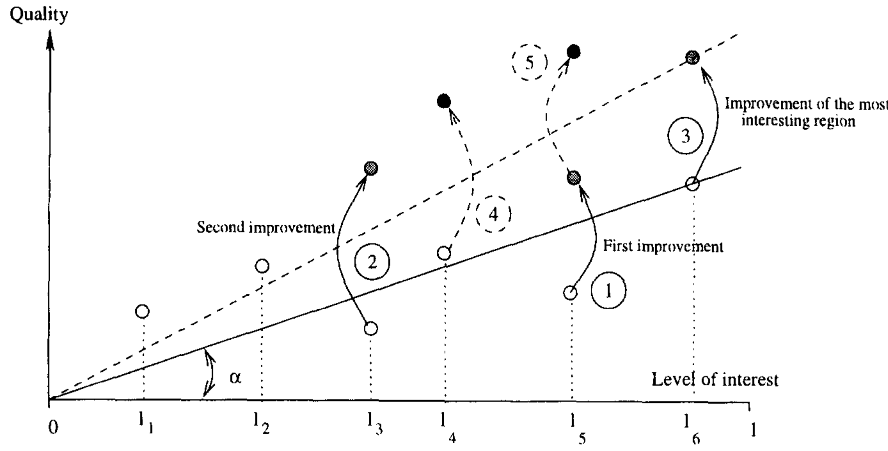


Fig. 6. Illustration of the allocation decision procedure under the assumption that the quality is directly proportional to the level of interest.

In the following, a practical allocation method is presented. It uses the previously exposed comparison with a straight line in order to avoid an exhaustive search. The way this practical and effective solution constitutes a particularly good alternative towards the optimization of (11) is discussed in Section 3.2.4.

3.2.3. *Practical solution*

Assuming first that the aim is to make the target quality $\hat{q}_{R_i}$ of a region R_i directly proportional to the level of interest of the region, the bit allocation scheme puts the representative points (l_{R_i}, q_{R_i}) of each region on a straight line passing by $(0,0)$ and having a slope as high as possible (under the constraint of B).

In order to avoid an exhaustive search, it is proposed to construct the solution step by step. It starts with an initialization step that encodes each region very coarsely, i.e. with the lowest quality, consuming the minimum amount of bits ($b_{\min, P_k}$). Then, if the bitrate allows it, it tries to refine the coding of the regions having an insufficient quality. At this stage, one has to choose the regions that is first improved. The way the decision is taken is illustrated in Fig. 6. A straight line, passing by $(0,0)$ and $(\max(l_R), q_{\max(l_R)})$ is drawn. The region having its representative point at the largest distance from the line is first improved. Each improvement results in a jump of the representative point on the graph. This jump corresponds to a discrete evolution of the coding parameters. Once all the points are above the straight line the point to be moved is the more interesting one. This process is run until there are no more bits available (limit B) or until the maximum quality has been reached for all the regions.

Assuming that the aim is to make the target quality linear with respect to the interest level, the allocation scheme implementation can be made more flexible. In Section 3.2.2, this is made through the introduction of paramater θ. Here, for implementation facilities, the constraint is expressed by a straight line passing by $(\max(l_R), q_{\max(l_R)})$ and whose slope α decreases as the obtained quality increases (function $A(q_{\max(l_R)})$).

Fig. 7 illustrates two different cases of slope evolution. These ones correspond to the requirements of users who do not have the same purposes. The first one prefers to keep a strong quality difference between interesting regions and non-interesting ones, while the second prefers to reach a uniform quality distribution as soon as the level of quality of the most interesting part of the scene is satisfying. This example illustrates how flexible the proposed method is with regards to users requirements.

Appendix A provides a C program implementing the regulation scheme.

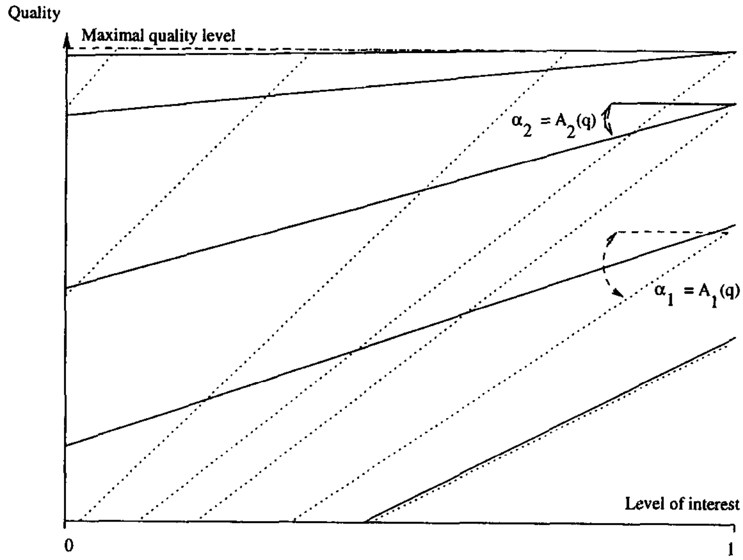


Fig. 7. Slope evolution for two different kinds of users. The first user (dotted line) prefers to keep a strong quality difference between regions of different interest while the second user (continuous line) prefers to have a more uniform quality distribution if the available bit-rate increases.

3.2.4. Practical solution and rate-distortion theory comparison

It exists a strong bound between the practical solution presented above and the solution of the maximization of the lagrangian cost function formulated through Eq. (11).

Concerning the optimization problem, the trends towards high quality and towards a linear relation between quality and level of interest is obvious and is expressed by the two terms of the lagrangian cost function.

In the practical method, as the allocated bit-rate is increasing, the couples (q_{R_i}, l_{R_i}) tend towards the required linear relation. The fact that $\Delta q_{R_i}/\Delta b_{R_i}$ decreases as q_{R_i} increases permits to finely adjust the points around the straight line. This is illustrated by the results of Section 4.2.1.

The trend towards a high global quality results from the fact that, among the regions that do not have a too high quality in regards with their interest, the improved region is always the lowest quality one. Once again, due to the decreasing behaviour of $\Delta q_{R_i}/\Delta b_{R_i}$, this choice will coincide with the one which would best improve the global quality.

Actually, the practical and optimal method provide solutions that are close to each other only if these solutions are effectively around the required straight line. This occurs when the global allocated bit-rate is sufficient. If it is not the case, the optimal solution is distinct from the practical one. Due to the fact that non-interesting regions are often less complex than interesting ones, the minimal bit-rate allows them to reach a rather high quality.

In this case, considering the optimal solution, coding the most interesting region with a higher quality highly improves the second term of the lagrangian cost function. Indeed, by increasing $q_{\max(l_R)}$, $\Theta(q_{\max(l_R)})$ is increased too and so, the part of the second term of the lagrangian cost due to regions of low interest decreases.

On the other hand, the practical method provides another solution. It ignores regions with low interest and high quality and distributes the available bit-rate among the other regions according to their level of interest. This permits to reach a solution which is better in accordance with the goal followed by the maximization of the lagrangian cost. The conclusion is that the behavior of the practical solution is more a quality than a weakness.

4. Results

4.1. Region classification

Figs. 9 and 10 present the results obtained for some MPEG4 sequences.

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 [30] divides the image into small regions whose edges include real objects edges. Oversegmentation inherent to this approach (Fig. 8(c)) 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, obtained results 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 (Fig. 8(b)). Another way to reduce oversegmentation is to make use of markers [13,23]. This approach, useful 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 worse 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 consists 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 (Fig. 8(d)).

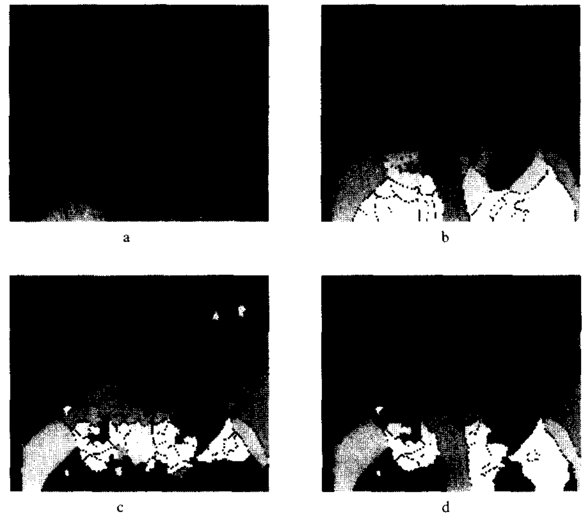


Fig. 8. Results of segmentation for 'Mother & Daughter': (a) original image, (b) segmentation with higher thresholds (daughter's head is merged with background), (c) oversegmentation (obtained with rather low threshold), (d) final result after merging.

In Figs. 9 and 10, 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.

Although the first results are promising, 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. Moreover, for the motion criterion, the definition of contrast does not seem to be the most judicious and thus needs to be improved.

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

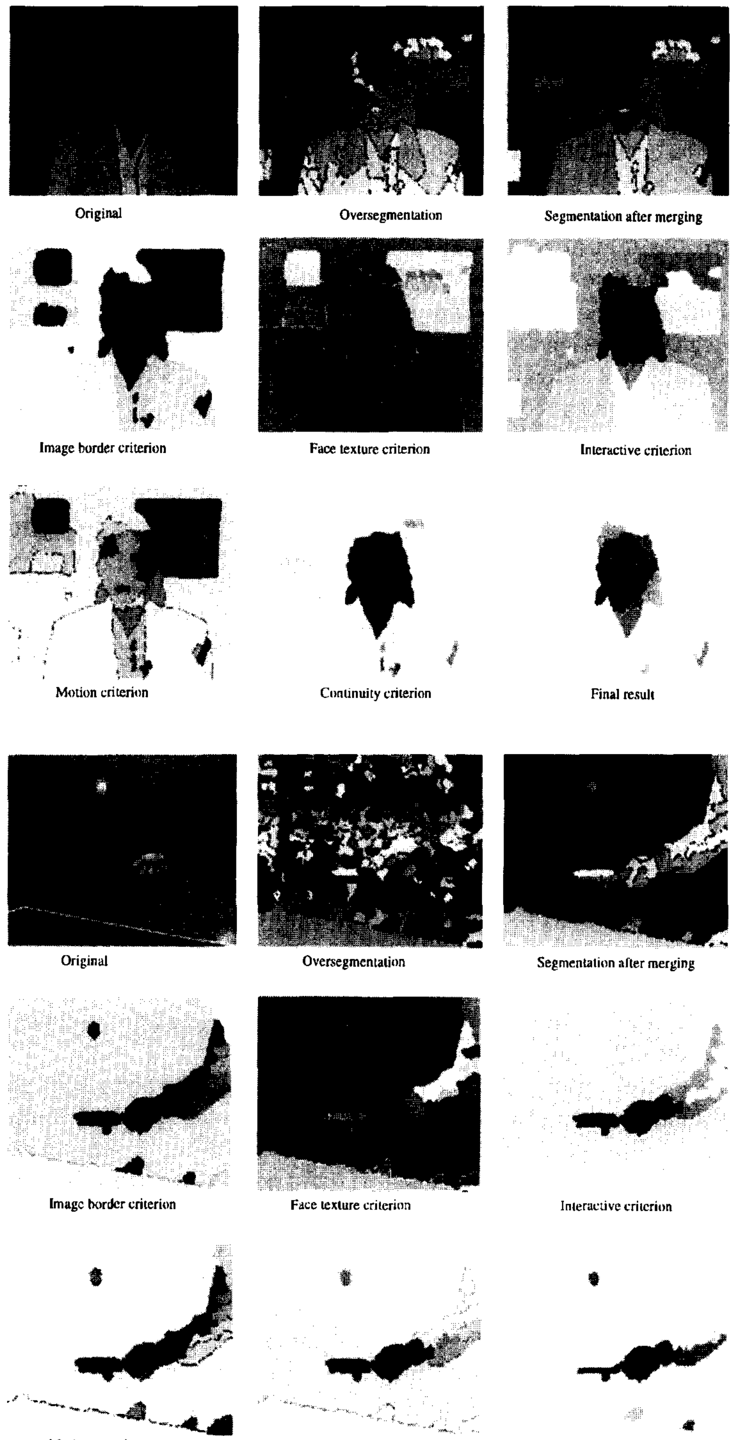


Fig. 9. Results of a first implementation. Sequences 'Akiyo' and 'Table Tennis'. The darkest regions are the most interesting ones. The results obtained for each criterion illustrate the defuzzification output obtained from the aggregation of the rules of the studied criterion.

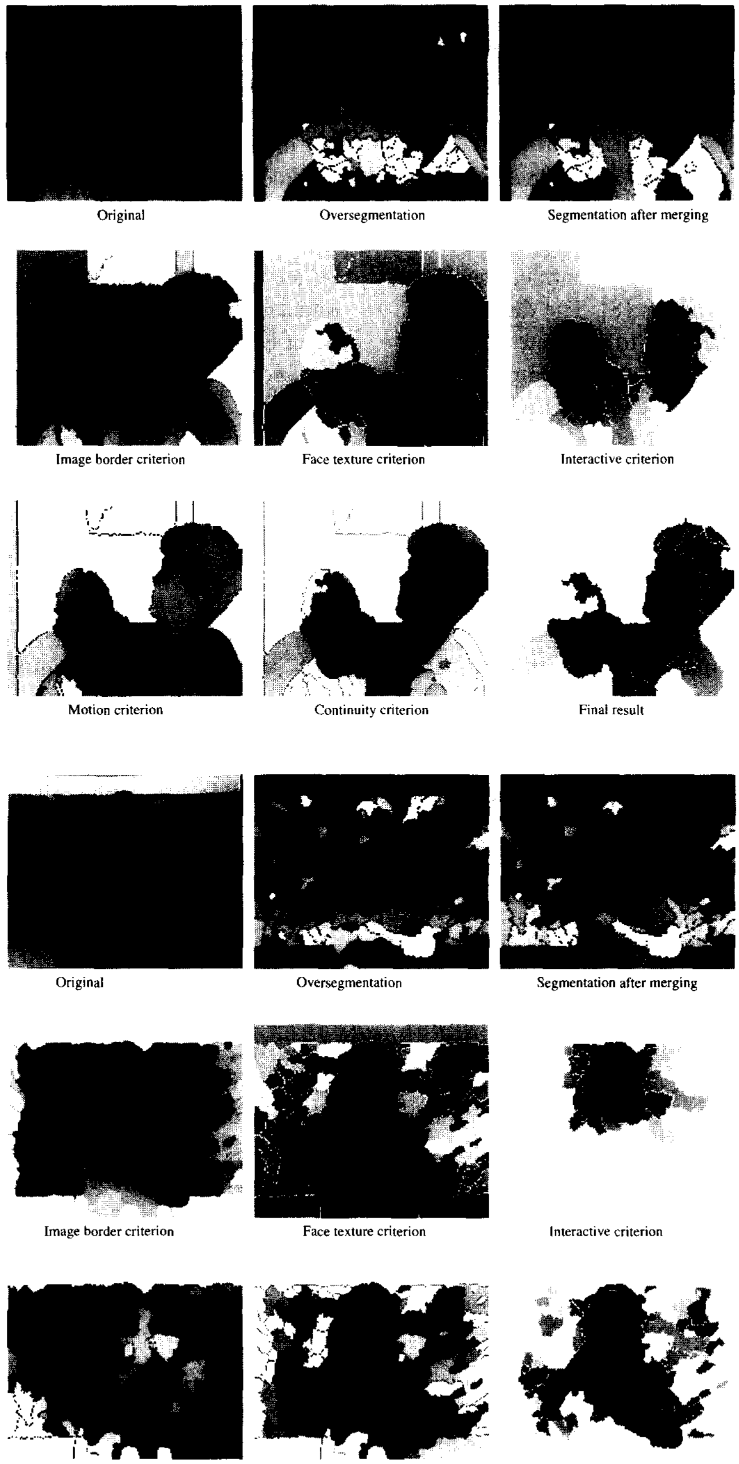


Fig. 10. Results of a first implementation. Sequences 'Mother & Daughter' and 'Silent'. The darkest regions are the most interesting ones. The results obtained for each criterion illustrate the defuzzification output obtained from the aggregation of the rules of the studied criterion.

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 [8]. 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 [14] and would probably be an interesting way to objectively match a level of interest to the training data.

4.2. *Preliminary results of subjective coding*

For the results presented in this section, a semi-automatic segmentation method has been used. This method, also based on the watershed algorithm and the use of markers [30, 13, 23], requires the user interaction to choose the markers and to force the merging of regions belonging to a same object. This permits to generate a semantic segmentation in accordance with the hypothesis of Section 2.

4.2.1. *Practical bit-rate allocation for intra-coding*

Figs. 11 and 12 illustrate the way the practical algorithm performs. Given a segmentation R and a level of interest for every region, the algorithm automatically divides up the bits. The quality criterion used here is the PSNR. The coding process just consists in a mere quantization of the grey level values with a variable step of 128, 64, 32, 16, 8 or 4. On the left of every picture is the situation of the regulation process and on the right the level of filling of the buffer.

4.2.2. *VM coder*

Results in this section have been generated thanks to the software implemention of the MPEG-4 Verification Model [17] provided by the MoMuSys ACTS project (software revision of August 96). The aim is to prove the efficiency of the proposed bit-rate regulation along a video sequence. For this purpose, a video sequence (Akiyo) has been partitioned into five Video Object Planes (VOP) (see Fig. 13(b)). Each VOP has received a level of interest from the FLS described in Section 2 (see Fig. 13(c)).

Given a bit-rate constraint (128 kbits/s), a frame rate (10 Hz) and a required slope α (= 10 dB on the interest range, see Section 3.2.3), the regulation scheme modifies the VOP's quantization parameter (QP) to reach a bit-rate balance corresponding to the VOP's interest level. The process involves a step by step method. The initial VOP's quantization parameter value is attributed acording to the method described in Section 4.2.1. Then, for each frame of the sequence, only the QP of one elected VOP is modified. This VOP is chosen using the method described in Section 3.2.3 and the quantization parameter is modified using a local bit-rate against average remaining bit-rate comparison method. Actually, this method compares the coding cost of the last frame that has been coded to the available remaining number of bits. If this cost is too high (low) to reach the bit-rate constraint, the QP is increased (decreased). The amplitude of the modification follows the Momusys recommendation as described in document Momusys/WG2-0040. This method compares the coding cost of the last frame that has been coded to the available remaining bit-rate. If this cost is too high (low) to reach the bit-rate constraint, the QP is increased (decreased).

Given the same bit-rate constraint, another regulation, using a common quantization parameter for each VOP, has been performed. At each time along the sequence, this parameter is also modified according to MoMuSys adjusting method.

Fig. 14(a) shows the QP evolution for both methods. In Fig. 15, each graph gives the PSNR to time relation for each VOP coded by the two strategies. It appears clearly that the proposed subjective regulation improves the quality of interesting area and decreases the quality of non-interesting ones. This behavior is confirmed by the total number of bits used for each VOP (see Fig. 14(b)).

Lastly, Fig. 16(a) presents the PSNR evolution for each VOPs. For the first intra-picture, in spite of the fact that they are coded with finer quantization parameter, VOP 2 and 3 do not reach a higher PSNR than VOP 4 and 5. It can be understood if one considers that the non-interesting regions reach a high PSNR level, even when coded with large quantization parameter, and that the available number of bits does not

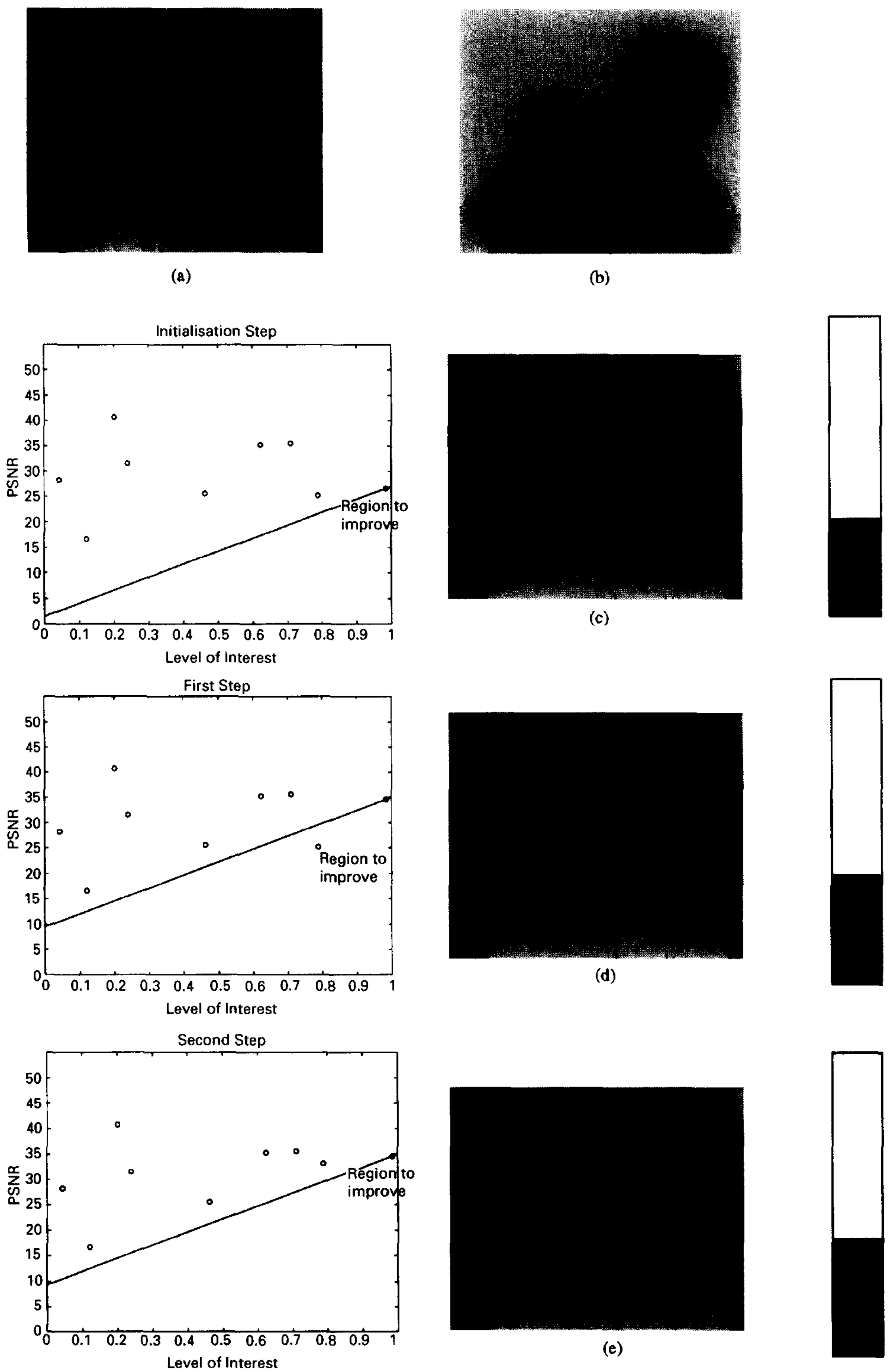


Fig. 11. Example of content-based bit-rate allocation according to subjective priority: (a) original image, (b) segmentation and related of interest (dark = interesting, light = less interesting), (c)–(e) first steps of the algorithm and associated pictures.

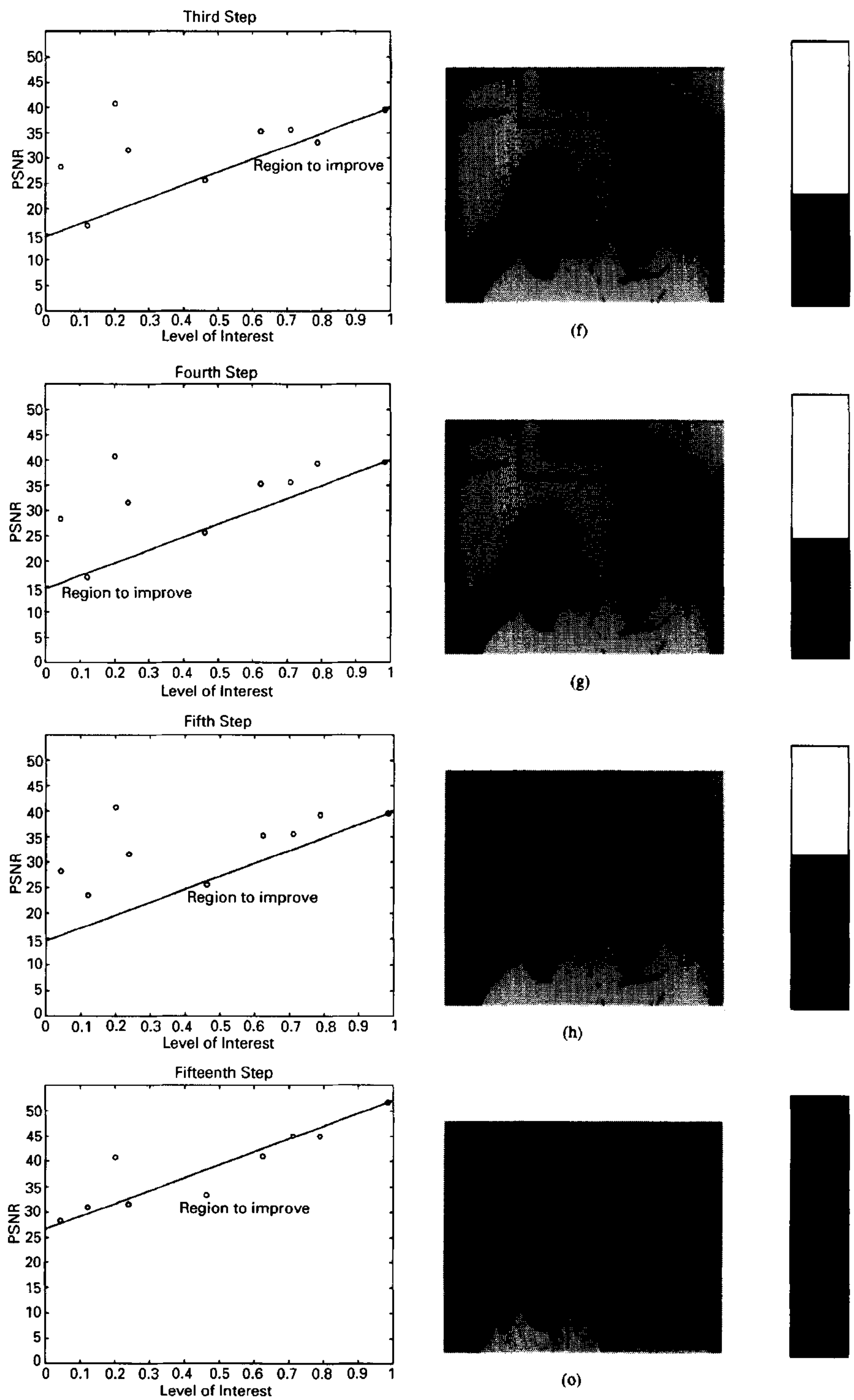
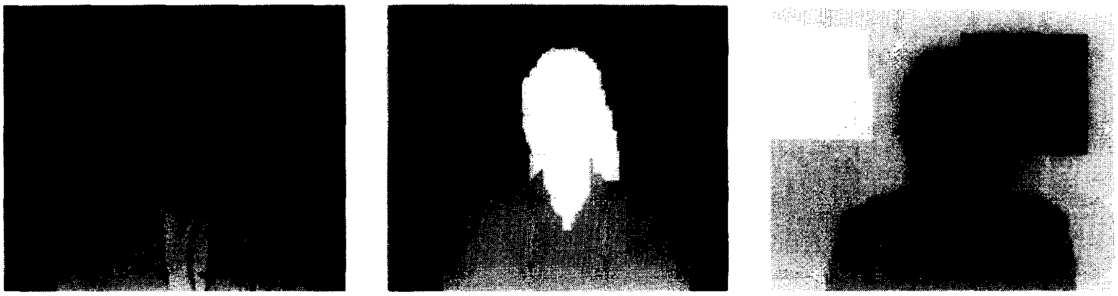
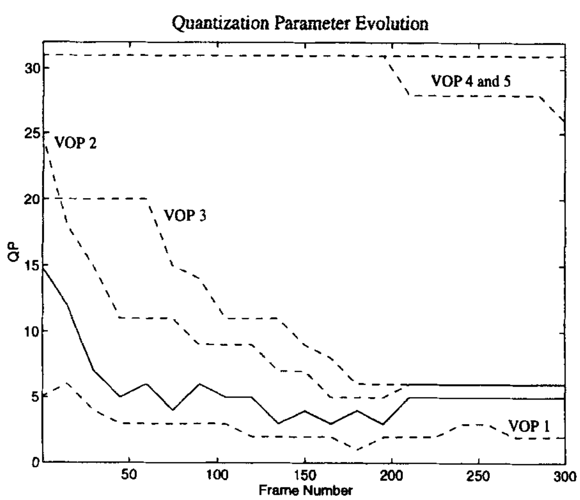


Fig. 12. Example of content-based bit-rate allocation according to subjective priority: (f)–(i) Various step of the algorithm and associated pictures.



(a) (b) (c)

Fig. 13. VOPs illustration for the first picture of Akiyo: (a) original picture, (b) VOPs (each grey level corresponds to an alpha-plane), (c) level of interest of the VOPs (darkness represents interest).



-------- = Proposed subjective VOPs regulation

———— = Classical frame regulation

(a)

	Subjective VOPs regulation	Classical frame regulation
VOP 1	736572	413275
VOP 2	269022	363037
VOP 3	48885	68656
VOP 4	43378	49407
VOP 5	176206	384191

(b)

Fig. 14. (a) Evolution of the quantization parameters for the two compared methods. (b) Total number of bits for coding each VOP.

allow interesting regions to reach a sufficient quality. Further in the sequence, as the remaining bit-rate allows it, the quality of the interesting regions is first increased. Conversely, in case of a common frame regulation scheme (see Fig. 16(b)), all PSNR curves have the same behavior.

Although the subjective quality improvement is effective and appears clearly when displaying the all video sequence, on still pictures, it is less obvious. This is due to the fact that the high bit-rate constraint permits to reach a high quality level for the all picture. At this level of quality, only small improvements appear on interesting areas. For this reason, pictures shown in Fig. 17 have been generated with a bit-rate constraint of 64 kbits/s. At this bit-rate, the subjective improvement appears clearly, even on still pictures.

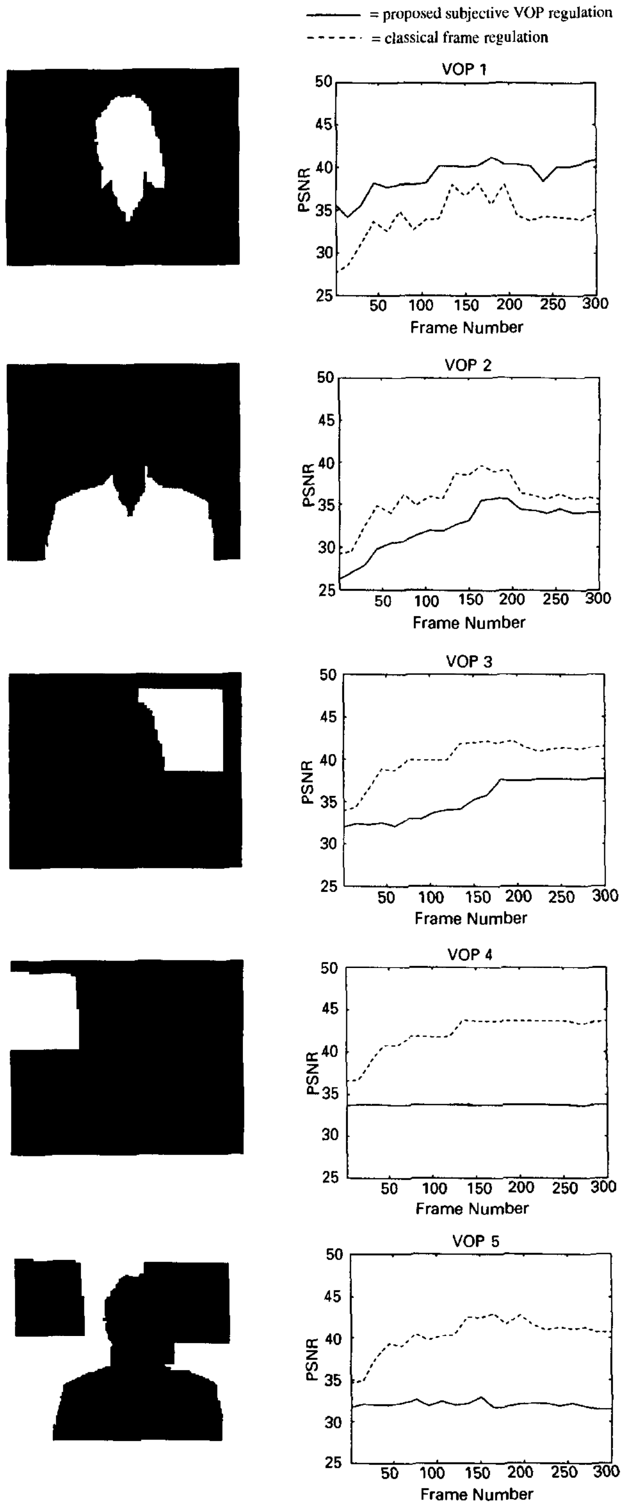
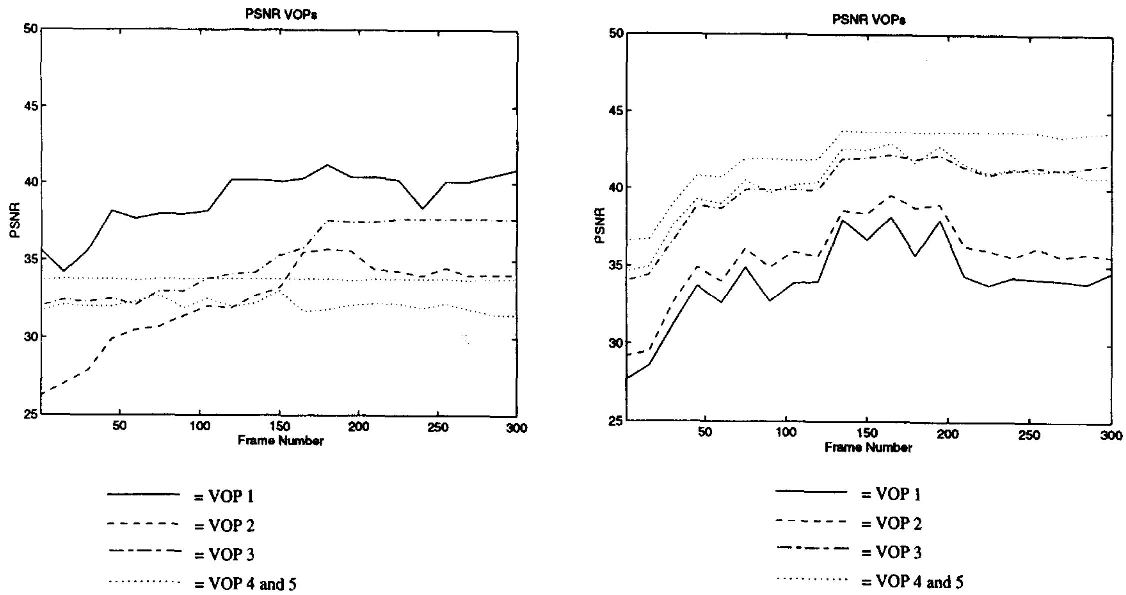


Fig. 15. Each graph presents the PSNR of a specific VOP as a function of time. Two coding process are compared: the proposed subjective allocation scheme and a classical frame allocation scheme using the same quantizing parameter for each VOP.



(a) (b)

Fig. 16. PSNR evolution: (a) proposed subjective allocation scheme, (b) frame regulation scheme.

5. Conclusion

This paper has presented a new way to perform bit-rate allocation in any video coder. Up to now, single variable regulation schemes, achieving homogeneous quality over the whole image, are generally used. Although they provide good results for high bit-rate coding, these schemes reach their limit once they are implemented in the framework of very-low bit-rate compression. On the contrary, this paper has proposed another way to allocate the bits, based on the assumption that the quality does not have to be constant on all the regions of the video sequence. A quality 'modulation' can be performed according to the fact that some parts of the images contribute actively in the information content while other parts do not contribute very much.

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, in order to allow a 'subjective video coding', 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. The output of this fuzzy logic procedure constitutes one of the crucial inputs of the bit-rate allocation algorithm.

Finally, the algorithm takes into account these subjective priorities in order to selectively distribute the available bit-rate. It is also noticeable that the algorithm is flexible to user's needs. Different users may indeed have different requirements: some prefer to watch at images with a uniform quality, even if this quality is somewhat low at very-low bit-rates; others would better like to maintain high quality differences between different parts of the image although the bandwidth is large enough to ensure an acceptable global quality. The flexibility and the adaptability of the scheme makes it able to satisfy both categories of users.

First results have demonstrated the reliability of the classification tool, even in the case of non-perfect segmentation. Nevertheless, the use of fuzzy logic allows to easily enhance the five chosen criteria and to add other rules based on the human cognitive understanding of the pictures. Other preliminary results proves the bit allocation scheme efficiency both for picture intra coding and for video sequence coding.

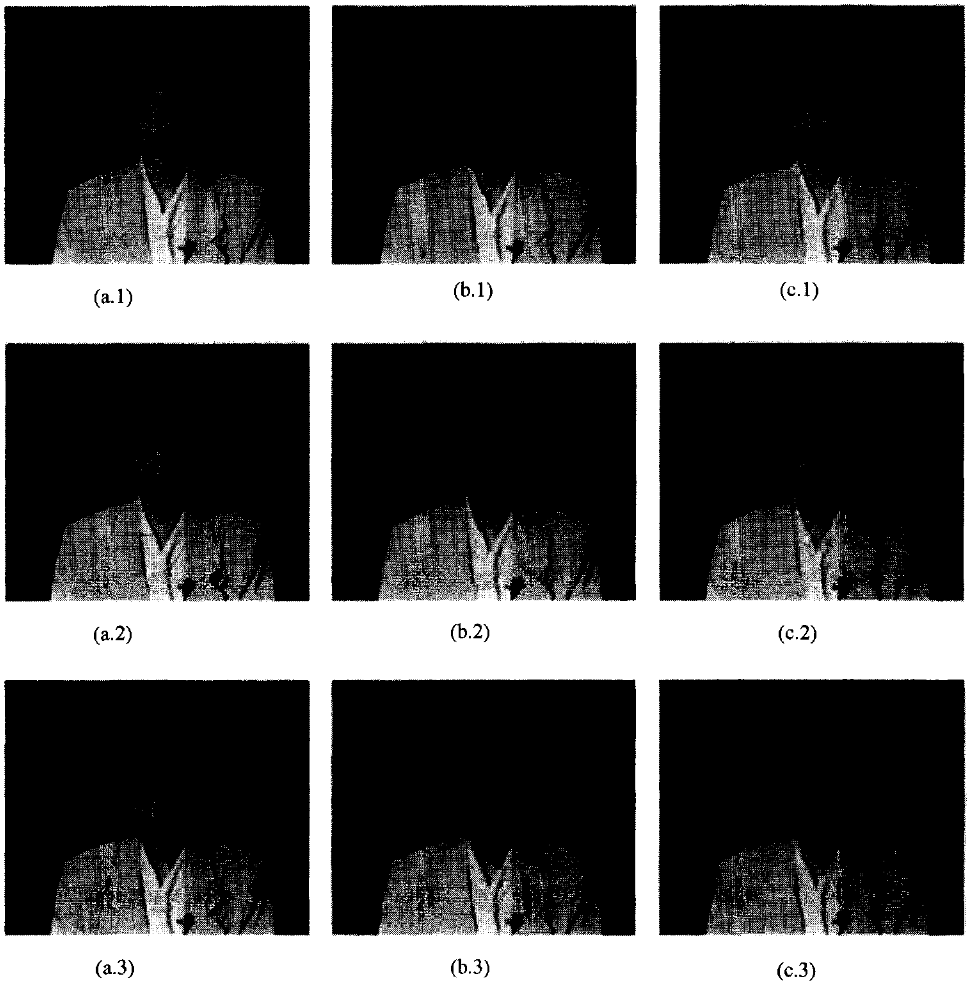


Fig. 17. Examples of coded pictures with a bit-rate constraint of 64 kbits/s (1=frame 1, 2=frame 148, 3=frame 300): (a) original, (b) classical frame regulation scheme, (c) proposed subjective regulation scheme.

Appendix A. C-like code of the bit allocation algorithm

The following notations and hypotheses are used for developing the algorithm of bit-rate allocation:

- Regions have been labeled according to their increasing level of interest.
- N denotes the maximum region label and is thus the label of the most interesting region.
- $\overline{p}_{\text{worst}}$ denotes the worst set of parameters (associated to $b_{\min,P_k}$). It is used to code each region in the coarsest way.
- q_{Max} denotes the best quality the coding algorithm will provide, with the finest available parameters.
- $b_{R_i} = \sum_{k|P_k \subset R_i} b_{P_k,\overline{p}_{P_k}}$. Let us remember here that, as stated in Section 3.1.2, the initial segmentation has been deformed so that the boundaries between regions are located on boundaries between partition elements.

- The variable 'Right' is the index of the most interesting region whose quality has not yet reached the best possible quality q_{Max}.
- The array distance[·] denotes the quality difference between the representative point of region and the reference straight line.
- The function Improve($\overline{p}_{P_k}$) modifies the parameters $\overline{p}_{P_k}$ in a minimal way in order to improve the resulting quality over the partition P_k.

```
/ * Initialisation * /
for(i = 1; i <= N; i ++)
    p̄_i = p̄_worst;
Right = N;
while(q_{R_Right} >= q_Max)
    Right = Right − 1;
/ * Iterative step * /
while(Σ_i b_{R_i} < B and Right > 0)
{
    α = A(q_{R_Right});
    / * Choice of the region to improve * /
    / * i_move = index of the region to work on * /
    i_move = Right;
    distance_{i_move} = distance[Right] = 0;
    for(i = 1; i < Right; i ++)
    {
        distance[i] = q_{R_i} − (q_{R_Right} − (l_{R_Right} − l_{R_i}) · tan(α));
        if(q_{R_i} < q_Max and distance[i] < distance_{i_move})
        {
            i_move = i;
            distance_{i_move} = distance[i];
        }
 }
    / * Quality improvement of the chosen region * /
    ∀k | P_k ⊂ R_{i_move},
        p̄_{P_k} = Improve(p̄_{P_k});
    / * Modification of the reference point if maximal quality has been reached * /
    if(i_move = Right and q_{R_Right} ≥ q_Max)
        Right = Right − 1;
}
```

References

[1] T. Berger, *Rate Distorsion Theory. A Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[2] J.C. Bezdek, "Partition structures: A tutorial", in: J.C. Bezdek, ed., *Analysis of Fuzzy Information*, CDC Press, Boca Raton, FL, 1987, Vol. 3, pp. 81–107.

[3] S. Comes, Les traitements perceptifs d'images numérisées, Ph.D. Thesis, Université Catholique de Louvain, June 1995.

[4] I. Corset et al., Technical description of sesame, Tech. Report, MPEG 95/408, November 1995.

[5] E. Cox, *The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems*, Academic Press, Cambridge, MA, AP Professional Edition, 1994.

[6] K. Hirota, Y. Arai and F. Hachisu, "Robot for recognition of moving objects and replacement of moving objects using fuzzy logic", *Proc. 2nd Fuzzy Systems Symp. of the Japan Chanter of IFSA*, 1985, pp. 15–22 (in Japanese).

[7] B. 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*, Prentice-Hall, University of Southern California, 1992, Chapter 1, pp. 29–32.

[8] B. Kosko, "Fuzzy associative memories", in: *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice-Hall, University of Southern California, 1992, Chapter 8, pp. 299–337.

[9] D. Le Gall, "MPEG: A video compression standard for multimedia applications", *Comm. ACM*, Vol. 34, No. 4, April 1991, pp. 46–58.

[10] B. Macq, M.P. Queluz and B. Simon, "Very low bit-rate image coding on adaptive multigrids", *Signal Processing: Image Communication*, Vol. 7, Nos. 4–6, November 1995, pp. 313–331.

[11] X. Marichal, An original approach towards very-low bit-rate image coding, *Revue HF*, Vol. 4, December 1995, pp. 29–46.

[12] J.M. Mendel, "Fuzzy logic systems for engineering: A tutorial", *Proc. IEEE*, Vol. 83, No. 3, March 1995, pp. 345–377.

[13] F. Meyer and S. Beucher, "Morphological segmentation", *J. Visual Commun. Image Represent.*, Vol. 1, No. 1, September 1990, pp. 21–46.

[14] 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, pp. 187–195, Grenoble, February 1996. CORESA 96 – Journées d'études et d'échanges: Compression et représentation des signaux audiovisuels.

[15] MORPHECO, "Morphological segmentation-based coding of image sequences", *COST 211ter European Workshop on New Techniques for Coding of Video Signals at Very Low Bitrates*, Hannover, December 1993, pp. 2.2.1–2.2.8.

[16] R. Morros, F. Marqués and P. Salembier, Video sequence segmentation based on rate-distortion theory, SPIE Visual Communications and Image Processing, Orlando, March 1996, Vol. 2727.

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

[18] H.G. Musmann, M. Hötter and J. Ostermann, "Object-oriented analysis–synthesis coding of moving images", *Signal Processing: Image Communication*, Vol. 1, No. 2, October 1989, pp. 117–138.

[19] F. Pereira, "MPEG4: A new challenge for the representation of audio-visual information", *Picture Coding Symp.*, Melbourne, March 1996, pp. 7–16.

[20] W.K. Pratt, Photometry and colorimetry, in: *Digital Image Processing*, Wiley Interscience, University of Southern California, 1978, Chapter 3, pp. 72–79.

[21] M.P. Queluz and B. Macq, *Signal-adapted Motion Compensation for Video Compression*, ISSES, URSI, Paris, September 1992.

[22] K. Ramchandran, A. Ortega and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders", *IEEE Trans. Image Process.*, Vol. 3, No. 5, September 1994, pp. 533–545.

[23] P. Salembier, "Morphological multiscale segmentation for image coding", *Signal Processing*, Vol. 38, No. 3, August 1994, pp. 359–386.

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

[25] Telecommunication Standardization Sector of ITU, Draft itu-t recommendation h.263, ITU Recommendations, July 1995.

[26] T. Terano, K. Asai and M. Sugeno, eds., "Applications in industry", in: *Applied Fuzzy Systems*, Academic Press, New York, 1994, Chapter 3, pp. 140–161.

[27] T. Terano, S. Masui, S. Kono and K. Yamamoto, "Recognition of crops by fuzzy logic", in: *2nd IFSA Congr.*, Tokyo, July 1987, pp. 374–377.

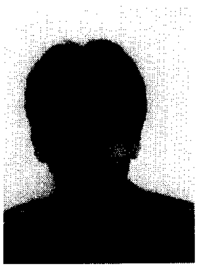
[28] M.M. Trivedi, "Analysis of aerial images using fuzzy clustering", in: J.C. Bezdek, ed., *Analysis of Fuzzy Information*, CDC Press, Boca Raton, FL, 1987, Vol. 3, pp. 133–151.

[29] C. van den Branden Lambrecht and O. Verscheure, *Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System*, SPIE Electronic Imaging: Science and Technology, San Jose, February 1996, Vol. 2668, pp. 450–461.

[30] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No. 6, June 1991, pp. 583–598.

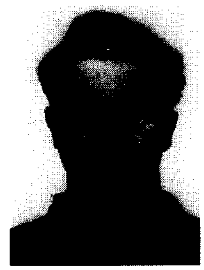
[31] G.K. Wallace, "The JPEG still picture compression standard", *Comm. ACM*, Vol. 34, No. 4, April 1991, pp. 30–44.

[32] R.R. Yager and D.P. Filev, "Template-based fuzzy systems modeling", *J. Intell. Fuzzy Systems*, Vol. 2, 1994, pp. 39–54.
[33] L.A. Zadeh, "Fuzzy sets", *Inform. Control*, Vol. 8, 1965, pp. 338–353.

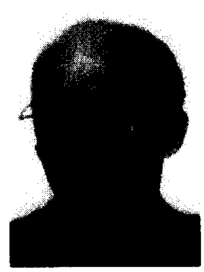


Christophe De Vleeschouwer was born in Namur, Belgium, in 1972. He received the Electrical Engineering degree from the Université Catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1995. From August to October 1995, he took part in the finalization of UCL MPEG-4's proposals. In October 1995, he received an F.N.R.S.- Alcatel Bell grant and he is now working on video coding and analysis as a Ph.D. student in the Laboratoire de Télécommunications et Télédétection of the UCL. His main interests concern video coding, image analysis and segmentation but also copyright protection.

Thierry Delmot was born in Uccle, Belgium, in 1970. He received the Electrical Engineering degree from the Université Catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in June 1994. In August 1994, he started working for Experteam S.A., a telecommunication software consulting SME. He joined the U.C.L. Laboratoire de Télécommunications et Télédétection in February 1995. His main mission is to represent BELGACOM during the MPEG-4 meetings and to participate in the standardization process. He took a major part in the finalization of the different UCL MPEG-4 proposals. His main interests concern image analysis, segmentation and semantic understanding (in the framework of very-low bit-rate video coding) on one hand, computer 'open networking' technologies and protocols on the other hand. T. Delmot received the 'Best Student Paper IEEE Award' during the DSP95 conference in Cyprus.



Xavier Marichal was born in Uccle, Belgium, in 1971. He received the Electrical Engineering degree from Université Catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in June 1994, after a four-month ERASMUS stay in Lisbon, Portugal, by Professor F. Pereira of the IST. In October 1994, he received an F.R.I.A-fellowship from the Belgian government and he is now working on video coding and analysis as a Ph.D. student in the Laboratoire de Télécommunications et Télédétection of the UCL. His main interests concern image analysis and motion estimation/compensation in the framework of very-low bit-rate video coding. He currently takes part in the European COST 211ter project.



Benoît Macq received the Ingénieur Civil Electricien and the Docteur en Sciences Appliquées degrees from the Université Catholique de Louvain (UCL), in 1984 and 1989, respectively. After his military service at the Ecole Royale Militaire, Brussels in 1984, he worked on telecommunication planning in the Tractionnel society in 1985, and on video coding in the Telecommunication Laboratory of the UCL from 1986 to 1990. From 1990 to 1991, he was with the Philips Research Laboratory Belgium. He has been a Senior Researcher of the Belgian NSF ('Chercheur Qualifié' du FNRS) and is now Professor and head of the Telecommunication Laboratory of the Université Catholique de Louvain. His current interests concern algorithms for image compression and video coding in broadhand digital networks, based on wavelets, neural networks, perceptual quantization, entropy coding and mathematical morphology, and multimedia access control and copyright protection for images.