Coding of Details in Very Low Bit-Rate Video Systems

Josep R. Casas and Luis Torres

***Abstract*—In this paper, the importance of including small image features at the initial levels of a progressive second generation video coding scheme is presented. It is shown that a number of meaningful small features called details should be coded, even at very low data bit-rates, in order to match their perceptual significance to the human visual system. We propose a method for extracting, perceptually selecting and coding of visual details in a video sequence using morphological techniques. Its application in the framework of a multiresolution segmentation-based coding algorithm yields better results than pure segmentation techniques at higher compression ratios, if the selection step fits some main subjective requirements.**

Details are extracted and coded separately from the region structure and included in the reconstructed images in a later stage. The fact of considering the local background of a given detail for its perceptual selection breaks the concept of "partition" in the segmentation scheme. As long as details are not considered as adjacent regions but isolated features spread over the image, "detail coding" can be seen as one step towards the so called feature-based video coding techniques.

I. Introduction

FIRST-GENERATION coding techniques [1] mainly attempt to diminish the statistical redundancy of image data. The coding schemes do not obey any perceptual image model and, when used for high compression ratios, the reconstructed images show annoying artifacts such as block effects, blurring and contour smoothing. The data representation used does not allow to go further in the perceptual selection of image features.

Second-generation coding techniques provide visual models that are able to handle severe selection of the visual information content. In second-generation schemes, the images to be coded are converted into a set of symbols according to some subjective image model [2]. Conversely to the former systems, the emphasis is put on the selection of these symbols before coding, rather than on the coding itself. The image model and the selection process are both designed to match the performance of the human visual system. For similar compression ratios, the better the matching, the higher the visual quality obtained after image reconstruction.

The *selection* of the visual features in the image plays a decisive role in any second generation coding scheme. The model used should make possible a ranking of these features close to the one yielded by subjective criteria (assuming that visual objects could be independently ranked). In this context, the usual tradeoff of the coding system, i.e., visual quality versus data rate, would be a matter of thresholding such a ranking, given the available data rate or, likewise, the compression ratio. The selection problem becomes critical at very low bit-rates, when significant losses in the reconstructed video sequence have to be assumed. For low compression ratios the selection is less critical, because most of the information discarded for coding is relatively of little significance or even redundant, but for high compression ratios, for instance about 50 to 100 for still monochrome images, the amount of discarded information will be as high as 98–99% [3]. Therefore, very little information of the original images will be kept after the coding process and, so, the balance between the selected and discarded visual features must be accurately observed.

In the framework of very low bit-rate video coding techniques, there is an increasing interest in second-generation compression techniques. These techniques eliminate redundant information within and between frames, taking advantage of the properties of the human visual system. In particular, segmentation-based coding methods try to describe the scenes in terms of uniformly textured regions surrounded by contours, in such a way that the regions correspond, as faithfully as possible, to the objects in the scene. The underlying image model takes into account probably the most significant feature picked up by the visual system: discontinuities or edge information [4]. Much effort is put on extracting and coding these "contours". The remaining features are seen as "textures" and coded roughly as some type of homogeneous distributions of grey level values. Due to the proximity of this image model to subjective perception, a better tradeoff quality-compression may be reached. Moreover, when the compression ratio is forced to achieve high values, the reconstructed video sequence shows a fairly graceful degradation of image quality.

Segmentation-based techniques have been successfully applied to still image coding [2]. The extension to video sequences involves the description of the scene in terms of regions in a 3D space, that is, as a three dimensional segmentation of the video sequence. The expression '3D space' refers in this case to both spatial dimensions plus the time domain considered as the third dimension.

Manuscript received November 3, 1993. This work was supported by the European Community within the RACE Project MORPHECO, and was recommended by Prof. Hans Georg Musmann.

The authors are with the Department of Signal Theory and Communications, ETSETB-Universitat Politècnica de Catalunya, P.O. Box 30002, 08071 Barcelona, Spain.

IEEE Log Number 9402531.

It is worthwhile to analyze the performance of segmentation-based schemes in a progressive coding framework [5]. In such systems, the number of regions selected for coding often depends on the target compression. If the compression is high, the homogeneity criterion is relaxed, so that the least "significant" regions are not segmented and they are merged into the most "similar" ones of their neighborhood. From this point of view, the significance of the regions and the similarity measure are none other than the homogeneity criterion for the segmentation algorithm. Nevertheless, as the compression increases the quality degradation of segmentation-based schemes is sometimes not so graceful as it should be. One may wonder if this is due either to some failure of the image model or to the selection process, i.e., the segmentation criteria. Improvements in both of them will be proposed in the present paper aiming for an increase in the coding rendition. Some new parameters will be computed in the original image for the selection step, whereas the concept of background will introduce a different viewpoint for the segmentation-based image model.

In particular, the selection process may be somehow improved in order to allow a better *perceptual selection.* Not only intrinsic features such as size, shape, absolute grey level or homogeneity of the amplitude distribution within the regions should be taken into account for the selection, but also extrinsic properties. We call "extrinsic" properties of a region those related to the local background. The *local background* of a visual feature, such as a region, is formed by the pixels in the immediate neighbourhood of the feature and the grey levels of the objects behind it, which are occluded by the feature. Extrinsic properties are the contrast of the feature over the background (relative grey level), the background grey level, its texture properties and the proximity to other dominant features in the image such as sharp transitions or highly textured areas. The incorporation of these properties into the coding scheme may lead to a quantitative step in second-generation image coding techniques.

As far as the image model is concerned, the fact of considering the local background of a region also implies a qualitative change. Not all the regions of a segmented image or video sequence may have a local background. Only the small regions will be seen as embedded in larger structures. The model of the segmented image as a complete partition of non-overlapping regions fails to some extent when regions are considered as superimposed image features occluding some parts of other regions: the background. The subjective visual ideas of foreground and background may break the segmentation scheme by making doubtful the restriction of "non-overlapping" regions for the image model. In that sense, the fact of considering background-foreground for some features in the image can be thought as a step towards feature-based image coding.

In this paper, small visual features called "details" are considered apart from the segmentation structure of adjacent non-overlapping regions. Details are extracted separately and perceptually ranked in order to achieve a good selection of the meaningful ones. Coding only the most significant details, improves the performance of segmentation-based schemes at very low bit-rates, i.e., the tradeoff visual rendition{/}bit cost is better than that obtained by pure segmentation techniques.

The paper is organized as follows. Next section presents a particular but well established segmentation-based video coding scheme, which will be used as reference for the application of the proposed technique. Section three is devoted to the description of the detail extraction technique, which is based on tools from mathematical morphology. To that extent, a brief introduction to morphological filters is given. In section four a solution is devised in order to obtain a perceptual ranking of the extracted details. Section five puts forward one method for the efficient coding of these features, according to the perceptual properties they have. In section six, results are shown for the detail coding technique when applied to the segmentation scheme described in section two. These results will be compared with the conventional segmentation performance of the same method. Finally, the last section gives some concluding remarks.

II. Morphological Segmentation-Based Video Coding Algorithm

The detail encoding technique will be presented in a progressive segmentation-based framework that has been originally developed for still image coding in [6] and extended to video sequences in [7]. The video coding algorithm is based on a three dimensional morphological segmentation scheme and on a contour-texture approach for the coding of the segmented sequence. An outline of the segmentation step and of the coding of contours and texture is provided below for completeness purposes. Only a short explanation is in order here for a better understanding of the interactions, performances and improvements of our method. For the interested reader, a more detailed explanation of the segmentation-based coding scheme can be found in [8].

A. Hierarchical Segmentation

The segmentation scheme is hierarchical in the sense that it leads to a set of segmented sequences ranging from coarse to fine. This feature is particularly attractive for progressive coding and transmission.

The selected video coding algorithm produces, in a first step, a coarse coded sequence with only a few regions. This first sequence can be coded with a very high compression ratio. Then, the successive levels of the algorithm improve the quality of the segmentation by introducing new regions so that the rough segmentation in the first level is refined in the next levels using more local information. As a result, the visual quality of the coded sequence improves at the expense of the compression ratio.

Fig. 1 gives an overview of the segmentation scheme. At each level of the hierarchy the same four basic steps are performed (Fig. 1(a)): simplification, feature extraction, decision and coding. These steps are carried out by means of morphological tools. In particular, the simplification step follows a size criterion while, for the decision step, a very efficient tool, the watershed algorithm, locates the precise contours of the regions. Once the contours have been found, they are kept through the remaining levels of the hierarchy.

The information to be coded in each level consists of new regions that are extracted from the current coding residue. The *coding residue* has been defined as the difference between the original image and the reconstructed image at the current level of the progressive hierarchy. A given hierarchical level in the segmentation process will have to deal with those components that have not been properly segmented and coded in previous levels and, therefore, can be found in the coding residue. In order to have information about these components, the sequence is actually coded (contour and texture coding) and the coding residue is computed and transmitted to the next level as an input image, as shown in Fig. 1b.

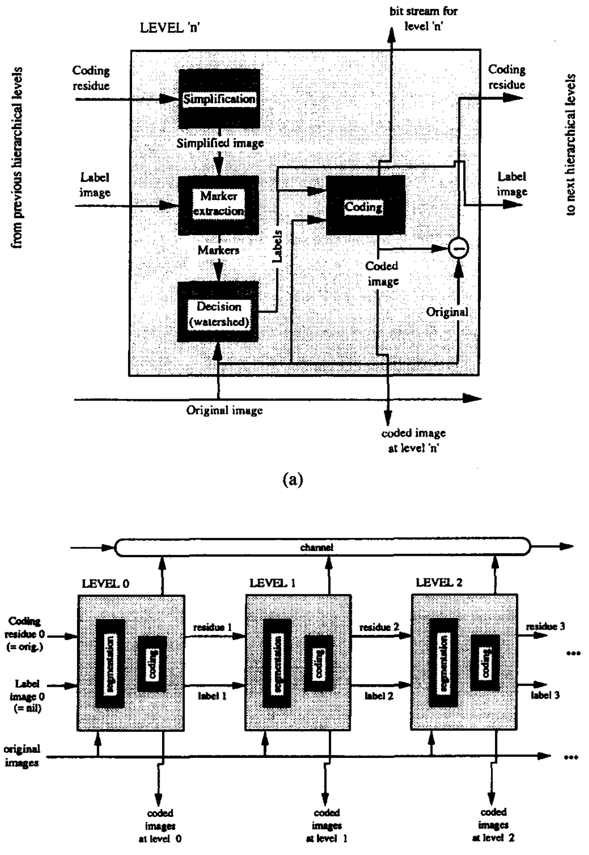
B. Contour-Texture Encoding

There are two different kinds of information to be taken into account for contour coding: shape and location information. Shape information refers to the form of each contour while location information deals with the position of each contour within the image. Contours are coded with the method proposed in [10]. This method starts by coding the spatial contour of the first frame in intra-mode by using the "chain code" technique [11] improved in [12] by the concept of "triple point". It leads to an average number of 1.3 bits per contour point (this figure includes both the contour description and the location of the starting points of each contour). For the following frames, the coding procedure makes a prediction of the contours of the future frame by using motion compensation, computes the contour prediction error, simplifies the error with morphological tools and transmits useful prediction errors. The motion information consists of one vector describing the translation of one region between two frames. In the average, the prediction error and the motion information can be coded with 0.4 bits per contour point. This contour coding technique is not lossless but the loss in visual quality is very small.

The inside of the segmented regions can be coded using some approximating function. In segmentation-based image coding techniques, polynomial approximations are usually employed. In this approach and in order to obtain very low bit rates, the region is approximated using the mean value within the entire volume of the region. Other approaches could be more useful but at the expenses of increasing the final bit rate. The mean value of the 3D region encoded with 8 bits has given good visual results for video sequences in QCIF format.

III. Detail Extraction

In order to address the problem of the extraction of significant details from digital images, a processing technique strongly related to the physical image structure is required. Such a technique, should deal with the "shapes" contained in the video signal. Furthermore, it should be able to infer the background lying underneath, which is occluded by the foreground shape. Mathematical morphology provides tools that give a good insight into the structure of the image for processing purposes.



(a)

(b)

Fig. 1. General morphological segmentation-based coding algorithm: (a) basic step; (b) hierarchical structure.

Mathematical morphology is a non-linear signal processing technique originated from the work of Matheron and Serra [13]. The morphological theory has sound mathematical foundations, but it can be used successfully with a very intuitive approach. Its original aim was to characterize physical properties by means of visual information. Although many signals combine additively, visual signals obey a very different way of composition. The physical world around us is generally made up of opaque objects and the nearest object occludes the one located behind it. Therefore, the first prerequisite that a visual-like transformation must fulfill is to preserve the existing relations of inclusion between every pair of objects, i.e., it must be increasing instead of linear[1] (the two notions are incompatible). Linear filtering techniques modify the object intensities and therefore the estimated location of

[1] In mathematical morphology, increasing transformations preserve the ordering relation defined in the working structure (lattice). The increasing property states that if an ordering relation holds for two input signals x_i and y_i, the same ordering is true for the output signals, i.e., if $x_i < y_i \Rightarrow \psi(x_i) < \psi(y_i)$.

In linear signal processing, the useful operations are also those preserving the working structure of the vectorial space and commuting with the fundamental laws (addition and scalar product). The resulting operation is the convolution: $\psi(\sum_i a_i X_i) = \sum_i a_i \psi(X_i)$.

their corresponding contours. Morphological filters examine the geometrical structure of images by probing their microstructure with certain elementary form, the structuring element, in the manner in which it fits into the image structure. Thus, the analysis is geometric in nature and derives quantitative measures from this point of view. Some basic definitions of morphological filters that will be used in the proposed algorithm for detail extraction are reviewed in this section. The reader is referred to [13] and [14] for further explanation.

A. Basic Morphological Filters

Morphological filters are defined as increasing and idempotent transformations. Being increasing, they preserve the ordering relation in the working space. The idempotency property limits the information loss by transforming in a single pass any original signal into an invariant signal. Morphological opening and closing are examples of morphological filters. They are based on the operations of erosion and dilation, which are defined below.

If 'x_i' and 'y_i' denote two signals defined in an N-dimensional space E^N, the erosion and dilation of 'x_i' by a window or flat structuring element 'B' of size 'n' are given by:

- **erosion**: $$y_i = \varepsilon_n(x_i) = \min_{k \in B}[x_{i+k}] \quad (1)$$
- **dilation**: $$y_i = \delta_n(x_i) = \max_{k \in B}[x_{i+k}] \quad (2)$$

Where 'i' indicates the location of the current sample and 'k' defines the distance to adjacent samples within the window. From a practical point of view, all the morphological filters in this study use a square structuring element and 'n' represents the square size. Morphological opening and closing of size 'n' are based on dilation and erosion definitions:

- **morphological opening**: $$y_i = \gamma_n(x_i) = \delta_n(\varepsilon_n(x_i)) \quad (3)$$
- **morphological closing**: $$y_i = \varphi_n(x_i) = \varepsilon_n(\delta_n(x_i)) \quad (4)$$

Opening and closing are dual filters, in the sense that the result of the closing is also the complement of the result of the opening applied to the complement of the original signal. For image signals, opening and closing allow to deal separately with bright and dark blobs in the image structure. As a result, the application of such filters can be seen as a special approach to the foreground-background concept for size and contrast features. The opening (resp. closing) simplifies the original signal by removing the small bright (resp. dark) components where the structuring element does not fit. Nevertheless, the contours of the large image components are often modified in order to fit the shape of the structuring element within the objects or within their background. In order to allow a perfect preservation of the contour information, a reconstruction process has to be used [14]. Its goal is to precisely restore the contour of the objects that have not been totally eliminated by the opening or the closing. Let us describe this reconstruction process.

Two dual reconstruction processes may be defined. After an opening, a positive reconstruction based on geodesic dilation has to be used, whereas in the case of closing, a negative reconstruction relying on geodesic erosion has to be used. Both geodesic dilation and erosion need an input signal 'x_i' and a reference signal 'r_i'. Their definitions are given for unitary size, that is the smallest window size in the digital case, which is normally taken as a symmetric 3 × 3-window in image processing.

- **geodesic dilation of size 1**: $$y_i = \delta^{(1)}(x_i, r_i) = \min[\delta_1(x_i), r_i] \quad (5)$$
- **geodesic erosion of size 1**: $$y_i = \varepsilon^{(1)}(x_i, r_i) = \max[\varepsilon_1(x_i), r_i] \quad (6)$$

These elementary geodesic dilation and erosion operators allow the introduction of the reconstruction processes, which are defined as iterated geodesic dilations or erosions. In practice, the unitary operations are iterated until idempotency, that is, until no change is observed in the output signal. Moreover, practical implementations of these functions can be done by very efficient techniques relying on waiting list structures that avoid any iterating process and lead to extremely fast algorithms [15].

- **positive reconstruction**: $$y_i = \gamma^{(\mathrm{rec})}(x_i, r_i) = \delta^{(1)}(\delta^{(1)}(\ldots \times \delta^{(1)}(x_i, r_i) \ldots, r_i), r_i) \quad (7)$$
- **negative reconstruction**: $$y_i = \varphi^{(\mathrm{rec})}(x_i, r_i) = \varepsilon(1)(\varepsilon^{(1)} \times (\ldots \varepsilon^{(1)}(x_i, r_i) \ldots, r_i), r_i) \quad (8)$$

In morphological image processing, the function to be rebuilt by the reconstruction process is usually seen as a "marker" image for significant bright or dark components of the reference image, whose locations are known but their exact shapes are not. Markers are, indeed, binary images identifying the presence of desired components. The original contours of these components will be found by means of the reconstruction process applied to the marker image taking the original image as the reference function. Finally, the opening and closing by reconstruction, used to preserve the edge information, are given by:

- **opening by reconstruction**: $$y_i = \gamma^{(\mathrm{rec})}(\varepsilon_n(x_i), x_i) \quad (9)$$
- **closing by reconstruction**: $$y_i = \varphi^{(\mathrm{rec})}(\delta_n(x_i), x_i) \quad (10)$$

Opening and closing by reconstruction are morphological filters as well, since they are increasing and idempotent, but their simplification effects in the filtered images are smaller than those of the morphological opening and closing. Large bright (resp. dark) objects which have not been completely eliminated by the morphological opening (resp. closing) are rebuilt to their original shape by the geodesic process so that their contours are preserved.

B. Special Morphological Filters for Detail Extraction

Although final results will be given for video sequences, for the sake of simplicity the discussion about the transformations involved in detail extraction will be illustrated for still images. Their performance is similar when applied to video sequences, considered as three-dimensional functions taking values within a horizontal-vertical-temporal domain. The structuring element for morphological operations is then a volumetric window

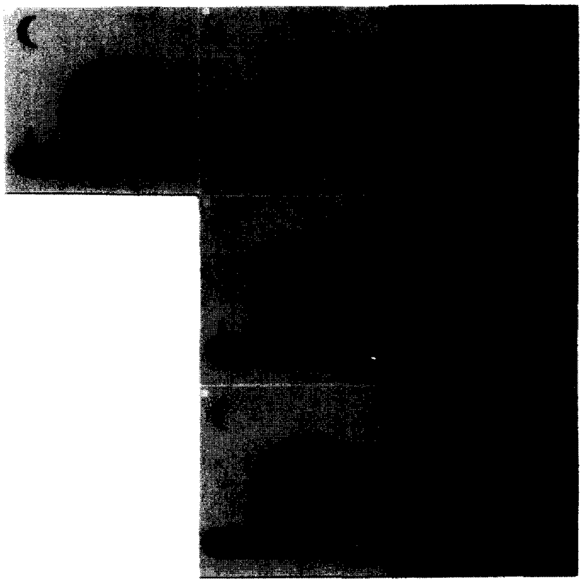


Fig. 2. "Top-hat" and "Post-it" on a synthetic image (a) original; (b) morphological opening-closing; (c) morphological top-hat; (d) opening-closing by rec.; (e) rec. top-hat; (f) post-it thinning; (g) details.

inside this 3D domain and the reconstruction process is done by geodesic propagation in three dimensions.

The *morphological top-hat* transform [16] is defined as the residue or difference between the identity operator and the opening: '$I - \gamma_n$' or, in the dual case, the closing and the identity: '$\varphi_n - I$'. They extract bright or dark contrasted components from the original image smaller than the structuring element, but also spurious components from the contours of other objects that have been modified by the morphological opening or closing. For the computation of the top-hat transform, one may also choose an opening or closing by reconstruction: '$I - \gamma^{(\mathrm{rec})}$' or '$\varphi^{(\mathrm{rec})} - I$'. With these filters, the shapes of the small objects are still visible on the filtered image after the reconstruction process. They simply get the same grey value as the neighbouring objects, becoming an extension of them. The residue of the opening or closing by reconstruction will be called in the sequel *reconstruction top-hat.*

A synthetic test image with some details is presented in Fig. 2(a). The application of morphological filters with and without reconstruction to the original image, namely opening-closing and opening-closing by reconstruction, produces the results shown in Fig. 2(b) and 2(d). The residual images, i.e., the results of the top-hat transform, are shown in Fig. 2(c) and 2(e). The morphological top-hat transform (Fig. 2(b)) extracts some features that do not correspond exactly to real objects in the original image. On the other hand the grey levels of the details extracted by the reconstruction top-hat (Fig. 2(d)) are dimmer than what they should be, because part of their amplitudes remains in the filtered image. Some improvements are needed for these filters in order to use them as detail extractors.

To overcome the drawbacks of both openings, a new morphological transformation called *post-it* has been proposed [17]. It is based on the selection of significant details from the reconstruction top-hat (Fig. 2(e)) and the computation of the true amplitude values for the selected details from the morphological top-hat (Fig. 2(c)). Let us denote these operators 'tht' and '$\mathrm{tht}^{(\mathrm{rec})}$'. They are applied to the original image 'f' or video sequence in order to obtain bright details. The dual operators, i.e. closing, will be used to obtain dark details.

- morphological top-hat: $$\mathrm{tht}(f) = f - \gamma_n(f) \qquad (11)$$
- reconstruction top-hat: $$\mathrm{tht}^{(\mathrm{rec})}(f) = f - \gamma^{(\mathrm{rec})}(\varepsilon_n(f), f) \qquad (12)$$

The image of bright details, 'det_w', is obtained by geodesic reconstruction of a marker image taking as reference the morphological top-hat. The marker image, 'mrk_w', indicates the location of significant details. It contains pixels of the reconstruction top-hat that are over a given contrast threshold 'λ'. By using the reconstruction top-hat to obtain the markers, artifacts due to contour smoothing, which do not appear in the reconstruction top-hat, are not rebuilt.

- marker for bright details: $$\mathrm{mrk}_w = \begin{cases} 255 & \text{if } \mathrm{tht}^{(\mathrm{rec})}(f) > \lambda \\ 0 & \text{otherwise} \end{cases} \qquad (13)$$
- image of bright details: $$\mathrm{det}_w = \gamma^{(\mathrm{rec})}(\mathrm{mrk}_w, \mathrm{tht}(f)) \qquad (14)$$

Details smaller than the structuring element and sufficiently contrasted to be "significant" are then obtained with their true contrast values over the background (Fig. 2(g). The smoothed image (Fig. 2(f)) is computed by subtracting the reconstructed details from the original image. Notice that at detail locations, the background of the bright (resp. dark) details has been filled up with the smaller (resp. larger) amplitude level of the neighborhood while the contours of the remaining components have been correctly preserved.

Fig. 3 shows the application of the detail extraction technique to the original 'cameraman' image. Please notice that the decomposition is quite intuitive from the visual point of view. The details of the smoothed images have been removed and their locations have been filled with values that fairly approach the perceptual notion of the background.

C. Modified "Post-It" for Extracting Details from the CodingResidue

In our approach, the original post-it transformation has been modified in order to obtain the details in a more suitable fashion for the posterior coding. The detail extraction technique will be applied in the progressive segmentation framework described in section 2. In this coding scheme, the information to be coded in the next level of the hierarchy consists of new regions that are extracted from the current coding residue. The detail extraction algorithm should be able to identify details from the coding residue as well, in order to improve the subjective quality of the coded segmentation.

(a) (b) (c)

Fig. 3. Detail extraction on still images (a) original "cameramen"; (b) extracted details; (c) smoothed image.

Therefore, the markers for these significant details are obtained from the coding residue instead of the original image. A contrast threshold '$\lambda = k_1\sigma$' is applied to the reconstruction top-hat computed on the residual image in order to select details. When an image feature does not present a contrast level in the residue larger than this threshold, it will not be marked for reconstruction, and so will not be extracted. So, since already coded details do not present high amplitude values in the coding residue, they will not be extracted for coding again. Moreover, the threshold could be made adaptive to the image and to the current hierarchical level by setting its value to some multiple of the variance of the coding residue.

If 'cod' stands for the current coded image, then the coding residue is '$\text{res} = f - \text{cod}$'. The modified post-it will follow the steps expressed in the sequel in order to get the marker image for bright details, 'mrk_w' (dark details will be obtained as usual by dual operators):

- morphological top-hat (original): $$\text{tht}(f) = f - \gamma_n(f) \qquad (15)$$
- reconstruction top-hat (residue): $$\text{tht}^{(\text{rec})}(\text{res}) = \text{res} - \gamma^{(\text{rec})} \times (\varepsilon_n(\text{res}), \text{res}) \qquad (16)$$
- marker for bright details : $$\text{mrk}_w = \begin{cases} 255 & \text{if } \text{tht}^{(\text{rec})}(\text{res}) > k_1\sigma \\ 0 & \text{otherwise} \end{cases} \qquad (17)$$

It is important to observe that the morphological top-hat operator is still applied to the original image in order to obtain true detail amplitudes, which could have not been preserved in the residue due to the coarse coding of the previous levels. Another modification with regard to the original post-it is the introduction of a second contrast threshold, '$k_2\sigma$'. While the first threshold is used for detail selection, the second one is applied to avoid noisy low-contrast blobs connected with selected details. This threshold is smaller than the previous one, i.e. $k_2 < k_1$, and is used to obtain a reference or mask image, 'msk_w', so that the reconstruction process does not propagate into connected spurious blobs:

- mask for the background: $$\text{msk}_w = \begin{cases} \text{tht}(f) & \text{if } \text{tht}^{(\text{rec})}(\text{res}) > k_2\sigma \\ 0 & \text{otherwise} \end{cases} \qquad (18)$$
- image of bright details: $$\text{det}_w = \gamma^{(\text{rec})}(\text{mrk}_w, \text{msk}_w) \qquad (19)$$
- smoothed image: $$\text{sm}_w = f - \text{det}_w \qquad (20)$$

The dual morphological transforms (top-hat of closing and closing by reconstruction) are applied on the smoothed image 'sm_w', in order to obtain the image of dark details 'det_b'. This sequence of operations yields the decomposition in detail image '$\text{det} = \text{det}_w - \text{det}_b$' and smoothed image '$\text{sm} = \text{sm}_w + \text{det}_b = f - (\text{det}_w - \text{det}_b)$'. Such a decomposition is parameterized by size and contrast; the size of the structuring element 'n' and the thresholds '$k_1\sigma$' and '$k_2\sigma$' being the control parameters.

Fig. 4 illustrates the application of the modified post-it transformation to the coding residue at a certain level of the segmentation hierarchy. Notice that those details which had already been fairly coded are not extracted now and that spurious low-contrasted components connected to the details are not reconstructed.

IV. Ranking and Selection of Details

For still images in the two-dimensional case, the detail image presents a set of connected components spread over a zero-valued background. For the three-dimensional space of video sequences, the details appear as narrow volumetric blobs, usually elongated in the time direction. Each connected component of the detail image is considered as a visual detail. In order to select the most meaningful ones, their supports are labeled individually in the so called "support image". Then the target is to get an ordered list of labels (ranking) as close as possible to the perceptual ordering that would be made by the human perception.

A first pre-selection is made when the first contrast threshold is set to '$k_1\sigma$' in the detail extraction step. This pre-selection reduces the number of connected components in the detail image, by discarding the least contrasted blobs. This makes

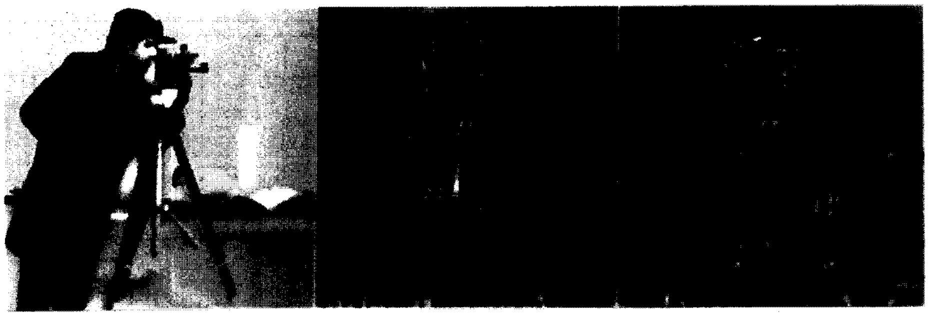


Fig. 4. Detail extraction from the coding residue (a) coarse segmentation (68 regions); (b) coding residue; (c) extracted details.

easier the ranking of the remaining details, because most of the least significant ones have already been discarded.

The criteria used for the selection step should be matched to the visual perception. Details are meaningful if they are contrasted over the background, even if they are small. Details of low contrast will be visible only if they have some significant size. Therefore, contrast and size will be taken as the most significant properties of each component in the detail image. Size is an intrinsic parameter, whereas contrast is an extrinsic one since it is related to the background where the detail is placed. In addition to contrast, some other extrinsic parameters will be considered. It has been noticed that, for the same contrast level, bright details are more visible in dark areas than in bright areas, while, on the contrary, dark details are more visible in bright areas than in dark areas. So, the grey level of the background has an influence on the perception of the small details and will be taken as an extrinsic parameter. Moreover, details located on smooth areas or flat regions are more visible than in highly textured areas or near to sharp transitions. Therefore, in addition to the contrast and the background grey level, a third extrinsic parameter will be considered: the spatial and temporal activity of the surrounding area.

In practice, details are obtained by the post-it transformation in a size-multiresolution scheme of three or four levels, with small but increasing structuring element sizes. The *size* parameter is given by the size-multiresolution level where the details have been extracted. The contrast selection will be stronger for small details than for the larger ones. The *contrast* parameter is defined as the average amplitude of the pixels in the detail image below each label of the support image. Contrast is taken as the main perceptual parameter in order to rank details of a given size. The other parameters are used as weighting factors applied to the contrast level. The *grey level of the local background* is the average amplitude over the set of pixels on the smoothed image below each label of the support image. The *background activity* is measured as the average of the morphological gradient [16] in the close neighborhood of each detail, i.e., in the area located only a few pixels away from the contours of the detail component. The contrast parameter of each detail is then weighted by the inverse of a linear combination of these magnitudes on the local background. If 'cn' is the contrast value for a given component and 'b_L' and 'b_A' are the background grey level and the activity measure, the ranking order will be given by the following empirical formula:

- **Ranking parameter**: $$\text{rnk} = \frac{\text{cn}}{\alpha \cdot b_L + \beta \cdot b_A} \qquad (21)$$

Further terms may be included in the linear combination, such as covariance measures or a priori information about where details should be. The coefficients of the linear combination may be varied according to the importance given to each perceptual factor. In our experiments, we have chosen values of 0.25 and 0.75 for the weights 'α' and 'β' respectively, which have yielded a ranking order quite close to a perceptual selection. Finally the weighted contrast values, 'rnk', are ranked in decreasing order and only some of the first details of the ranked list are selected for coding, depending on the available bit-rate. The support image for some of the details extracted in Fig. 4 is shown in Fig. 5(a), with the amplitude levels set to the contrast parameter 'cn' for each detail. Fig. 5(b) presents the background weighting factors; the more weight is applied to the contrast parameter, the brighter its location appears in this image. The top sixty details of the ranked list are shown in Fig. 5c and they have been included in the coded segmentation in Fig. 5d.

V. Coding of Selected Details

Not all the parameters of the details that have been selected for coding are equally important. Some of them must be coded more carefully than others. Three properties are considered for coding: amplitude, shape and location of details. They have been ordered according to their perceptual significance:

- amplitude: as the human visual system is not very sensitive to amplitude variations of the spatial and temporal high frequency components from which small details are mainly formed, the least important magnitude is the exact grey-level of the detail pixels, and the amplitude variations within same details may be simply discarded.
- shape: details have been obtained as small shapes. Slight variations of their contours are usually beyond visual resolution. If there is any advantage in simplifying the shape of any detail by adding or removing some pixels, it can be done without much trouble because a slight

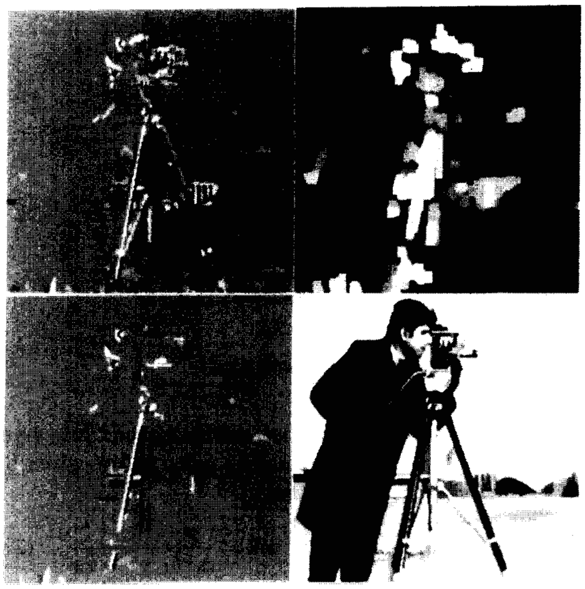


Fig. 5. Ranking of the extracted details (a) contrast parameter 'cn'; (b) background weighting: $(\alpha \cdot b_L + \beta \cdot b_A)^{-1}$, c) selected details, d) coded segmentation of Fig. 4(a) with details.

smoothing of some pixels in the detail contour will be hardly visible.

- location: is the most important parameter of a detail component and must be carefully coded. Small displacements of detail locations are easily perceived by the observer as deviations from regular positions or from positions which are a priori known.

In order to take advantage of the perception of the previous properties, a method has been designed that efficiently codes the exact position of each connected component and pays little attention to its amplitude and shape. Details are coded as small isolated regions of constant grey level. The average amplitude of the pixels within the detail support in the original image is taken as the detail amplitude. Amplitude values for selected details are coded with eight bits and put in scanning order in a buffer of grey levels whose length is equivalent to the number of details that have been selected. The shape and location of these details are obtained from the support image and coded using run-length coding techniques.

Multidimensional run-length coding, i.e., Relative Element Address Designate (READ) coding [18], is used to skip the background spaces and benefit from the connectivity of the small components for coding the support images. Run-length techniques have proven to be efficient in coding text and two-level graphics for the facsimile standard (CCITT group III). They are widely used in first generation image coding schemes for coding the locations of isolated high frequency coefficients in transform coding techniques. In [19], run-length coding has been used for feature coding in a second generation coding scheme. In our approach, the aim of the coding algorithm—very low bit-rate coding of video images—is quite different from those of facsimile or first generation techniques, but the structure of the support image (Fig. 5(a)) is quite suitable for the application of run-length coding. Besides, the temporal correlation of image details can be exploited by taking the runs in the temporal direction—3D run-length coding—and thus the efficiency is increased.

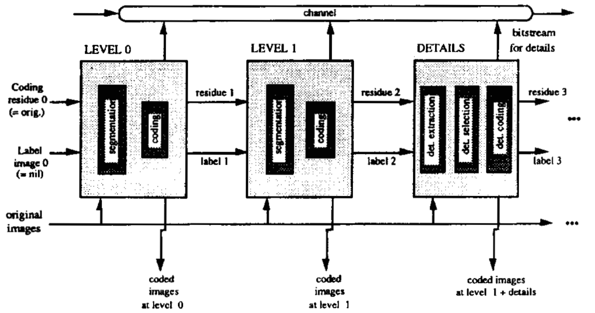


Fig. 6. Improved morphological segmentation-based coding scheme. The detail extraction step has been included at level 3.

The input binary image for the run-length algorithm is generated from the support image by binarization. To prevent binary supports of dark and bright details from being connected, some pixels of the contour of the details—detected as contact points—are removed. However, the probability for details of opposite contrast of being connected is rather small in most images, and so the shapes of some details may be slightly changed without noticeable artifacts. Differential runs are then obtained for the relative addressing of each detail and for relative variations of its shape. In the video sequence, the reference used for the coding of the current line is taken from the previous frame, in order to take advantage of the temporal redundancy.

Run-length can over-perform contour coding techniques like chain-code [11] when the details are smaller enough to have more contour points than inside points. When region contours are not connected to each other, the coding of the coordinates of the initial point for each contour severely penalizes chain-code techniques. Moreover, run-length coding allows progressive encoding if the details of the support image are distributed in different images according to their significance and encoded at different levels of a progressive hierarchy. Since there will be less and larger runs in each of these images, only a small load for the final bit-rate will be added.

Finally, an arithmetic coder is applied to both the buffer of grey-levels and the output of the run-length coder, in order to get actual bit-rates for the coded images.

VI. Results

In this section, we will present a comparison between the segmentation-based approach to video coding [8] at two different levels of the progressive transmission scheme and the introduction of detail coding for the small regions at the higher level, as it is shown in Fig. 6. We have applied this technique to black and white images, but the extension to color images is straightforward with a very small increase in the compression ratio.

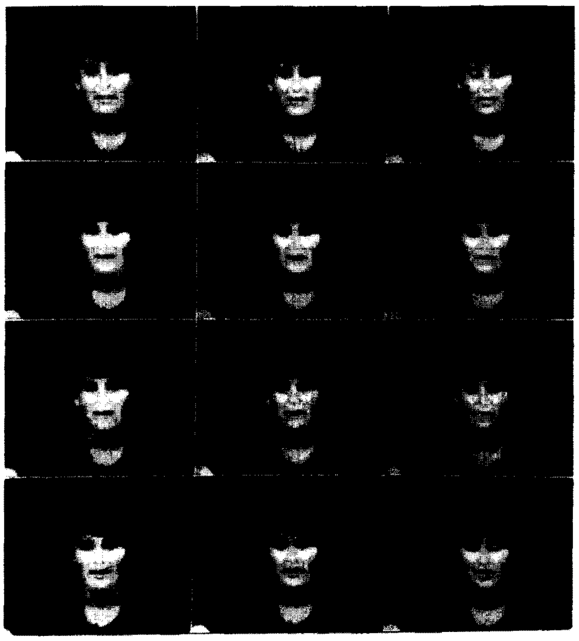


Fig. 7. Comparision of techniques: (a) three consecutive frames of the original sequence (first row); (b) first level of segmentation, 3 regions, 590 bits/frame (second row); (c) first and second levels of segmentation, 43 regions, 1036bits/frame (third row); (d) first level (from b) and detail coding, 23 regions + 18 details, 1010 bits/frame (fourth row).

Three consecutive frames from an original black and white video sequence in QCIF format the well known "Miss America" are shown in Fig. 7(a). The two rows below (Figs. 7(b) and 7(c)) present the coded images at two different levels of the progressive segmentation hierarchy, where respectively 23 and 43 (23+20) volumetric regions have been segmented. The number of bits used for the encoding of each frame (excluding the initial one) is in average 590 bits for the first level and 1036 (590 + 446) bits for the first and second levels. These figures include both texture and contour information. With a frame rate of 10 Hz, it would lead to final bit-rates of 5.9 and 10.4 Kbit/s respectively.

Although the bit-rate is almost doubled from the first coded segmentation to the second one, it can be noticed that the subjective quality improvement does not seem to be in accordance with a doubling of the cost. The target is, for a similar bit-rate, to improve the visual quality of the coded sequence at the second level, using the detail extraction, selection and coding technique described in this paper. As the segmentation scheme follows a size multiresolution criterion, we assume that all the significant largest regions have already been segmented and, instead of performing a new segmentation step from the lower level (Fig. 7(b)), we will simply add some small and contrasted components (meaningful details) in order to obtain a new version of the second progressive level. The available bit-rate of detail coding is the difference between the final rates of the two levels of the coded segmentation hierarchy, that is, 10.4 − 5.9 = 4.5 Kbit/s (exactly 446 bits per frame).

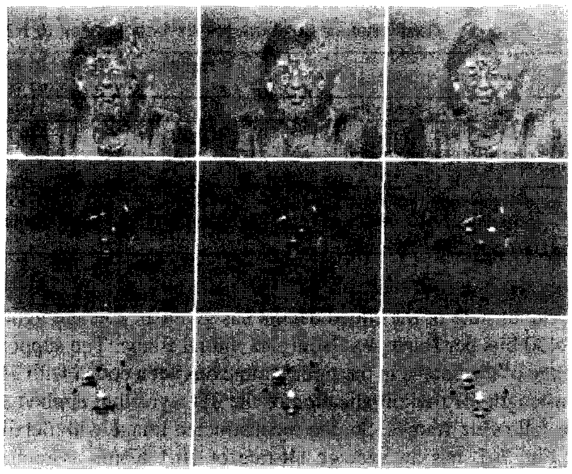


Fig. 8. Detail extraction from a video sequence: a) coding residue (first row), b) extracted details (second row), c) coded details (third row) Note: zero error is presented by middle grey.

In Fig. 8a, the current coding residue at the input of the second level of the segmentation algorithm is shown. This residual sequence has been computed as the difference between the original sequence (Fig. 7(a)) and the coded sequence at the previous level (Fig. 7(b)). Due to the fact that this sequence ranges from positive to negative error values, a constant value of 128 has been added in order to display the error sequence. Thus, when the reconstruction error is zero, the pixel is presented with an average grey level (128) in the displayed images.

The detail extraction technique is applied on this coding residue (17) in order to mark the location of meaningful details, whose exact contours and amplitudes are extracted from the original sequence (15) and (16). The result of the detail extraction step is shown in Fig. 8(b). Notice that the dark and bright components obtained are spread over a zero valued background and that the amplitude distribution within each component is not constant. The extracted details are ranked according to the 'rnk' parameter (21) and the first details in the ranked list are selected for coding. Only 10 bright details (out of 22) and 8 dark details (out of 18) have been selected, in order to accomplish the imposed limit of 4.5 Kbit/s for the bit-rate. The support image will thus contain 18 volumetric components for 3D run-length coding. The detail amplitudes are also coded using for the mean value within each 3D region encoded with 8 bits. The coded details are presented in Fig. 8(c). In order to make visible the darkest ones, the background has also been set to 128 for the displayed sequence.

The coded details have been included in the first level sequence of Fig. 7(b) by a simple logical "or" operation, that is, where the support image is not zero, the reconstructed image gets the value of the coded detail (Fig. 8(c)), otherwise the value of the image from the previous level (Fig. 7(b)) is kept. The reconstructed image including the 23 regions from the first segmentation level and the 18 coded details is shown in Fig. 7(d). Please notice the perceptual significance of the details that have been selected for coding.

TABLE I
COMPARISON OF BIT-RATES FOR A FRAME RATE OF 10 HZ. FIGURES WITH '+' ARE PARTIAL RATES FOR THE GIVEN LEVEL

	# regions	**# bits contour** (per frame)	**# bits texture** (per frame)	**# bits per frame**	**bit-rate** **(10 Hz)**
1st segmentation level (fig. 7b)	23 reg.	550bit/fr	40 bit/fr.	590 bit/fr.	**5.9** Kbit/s
2nd seg. level...	+20 reg.	+416 bit/fr.	+30 bit/fr.	+446 bit/fr.	+4.5 Kbit/s
1st + 2nd segmentation levels (fig. 7c)	43 reg.	966 bit/fr.	70 bit/fr.	1036 bit/fr.	**10.4** Kbit/s
details only (fig. 8c)...	+18 det.	+391 bit/fr.	+29 bit/fr.	+420 bit/fr.	+4.2 Kbit/s
1st seg. level + details (fig.7d)	23 + 18	941 bit/fr.	69 bit/fr.	1010 bit/fr.	**10.1** Kbit/s

The bit-rate of the coded details has given an average figure of 420 bits per frame, excluding the initial frame. The support image of the details in the initial frame has been coded with 2D run-length for the initialization of the 3D algorithm applied to the following frames. So, the reconstructed image with details of Fig. 7(d) has an average bit-rate of 10.1 Kbit/s assuming a frame-rate of 10 frames per second. Table I gives further information about these figures. The number of contour bits indicates, for both segmentation levels, the actual bits used by the 3D contour coding algorithm [10] after entropy coding (also excluding the first frame for initialization). In the case of detail coding, the figure in the same column represents the data-rate given by the 3D run-length coder for the support sequence after arithmetic coding.

It is worthwhile to observe the visual quality improvement that, for a similar bit-rate, represents the inclusion of the proposed technique for selection and coding of details. The technique has also been applied to other video sequences with the same results, what proves the goodness of the proposed algorithm. Numerical results obtained for video coding at very low bit-rates confirm the importance of the visual details for the quality of the coded image. These results can be improved by means of motion compensation techniques applied to the 3D run-length method used for detail coding. If instead of the previous frame, a motion compensated prediction of the current frame is used as reference, the temporal redundancy will be larger. The coded temporal runs will allow greater entropy compression, although a motion vector should be transmitted for each detail.

VII. CONCLUSION

A morphological-based technique for encoding the meaningful details of video signals has been presented. The perceptual significance of the details from a subjective point of view has been discussed along with the parameters involved in their perception. The details must be taken into account in second generation video coding techniques, when a compromise has to be found between bit-rate and visual quality of the encoded image. The significance of details is even greater in a progressive coding framework for very low bit-rate video coding schemes. Although the coding technique presented in this paper has been shown on a segmentation-based scheme, it has been applied as well to linear coding techniques [20] such as the linear pyramid decomposition, with similar results.

It has been shown that mathematical morphology provides suitable operators for the extraction of small video features. To this end, an efficient algorithm has been described. A perceptually matched ranking method has been presented for the selection of the most significant of the extracted details. Some perceptual parameters such as size, contrast, background contents are involved in the selection process. Finally, the selected details are coded according to their perceptual significance by means of a purpose-designed coding algorithm.

The results obtained in this work prove that both the selection process and the image model can be improved in segmentation-based techniques in order to increase the coding rendition. Details somehow break the model of non-overlapping regions used in segmentation-based techniques. They are coded independently and included into the reconstructed image in a later stage. Moreover, the concept of background considered for the selection step does not fit the partition structure of a segmented image, because details are seen as objects occluding the regions located underneath. From this point of view, detail coding may be considered as a texture coding method, although the contours of the selected details must also be coded. Some other new features in the image model, such as open contours, are under investigation. The aim is to find a proper matching of the coded features in the image model to the visual features in real images, by relaxing some restrictions in the segmented image model. This approach may lead in the future to a better performance in both visual rendition and bit-rate of segmentation-based video coding techniques.

ACKNOWLEDGMENT

The authors would like to acknowledge the invaluable help offered by P. Salembier and M. Pardas. Without their suggestions and without their contribution of the 3D segmentation algorithm this work would not have been possible. The comments and suggestions of the anonymous reviewers are also highly appreciated.

REFERENCES

[1] A. N. Netravali and J. O. Limb, "Picture coding: a review" *Proc. IEEE*, vol. 68, no. 3, pp. 366–406, Mar. 1980.
[2] M. Kunt, A. Ikonomopoulos and M. Kocher, "Second generation image coding techniques" *Proc. IEEE*, vol. 73, no. 4, pp. 549–574, Apr. 1985.
[3] W. E. Glenn, "Digital image compression based on visual perception and scene properties," *SMPTE Journal*, pp. 392–397, May 1993.
[4] T. Cornsweet, "Visual Perception," *Academic Press*, 1970.
[5] K-H. Tzou, "Progressive image transmission: a review and comparison of techniques," *Optical Engineering*, vol. 26, pp. 581–589, Jul. 1987.

[6] P. Salembier and J. Serra, "Morphological multiscale segmentation of images," *Proc. of SPIE Visual Comm. and Image Processing*, vol. 1818. Boston, MA, 1992.

[7] P. Salembier, L. Torres, M. Pardas, F. Marques, P. Hierro and A. Gasull, "Morphological segmentation-based coding of image sequences," *IEEE Europ. Conf. on Circuits Theory and Design*, Davos, Switzerland, Aug. 30–Sept. 3, 1993.

[8] P. Salembier and M. Pardàs. "Hierarchical morphological segmentation for image sequence coding," *IEEE Trans. Image Processing, Special Issue on Video Sequence Compression*, July 1994.

[9] F. Meyer and S. Beucher, "Morphological segmentation," *Journal of Visual Communications & Image Representation*, vol. 1, no. 1, pp. 21–46, 1990.

[10] C. Gu. and M. Kunt. "Contour image sequence coding by motion compensation and Morphological Filtering," *Proceedings of the International Workshop on Coding Technologies for Very-Low Bit-Rate Video*, Colchester, U.K., 7–8 Apr., 1994.

[11] H. Freeman, "On the encoding of arbitrary geometric configurations," *IRE Trans. Electron. Comp.*, vol. EC-10, pp. 260–268, Jun. 1961.

[12] F. Marqués, J. Sauleda and A. Gasull. "Shape and location coding for contour images," *in Proc. of the 1993 Picture Coding Symposium*, p. 18.6, Lausanne, Switzerland, 1993.

[13] J. Serra, *Image Analysis and Mathematical Morphology,* Academic Press, New York 1982.

[14] J. Serra, *Image Analysis and Mathematical Morphology, vol. 2: Theoretical advances*, Academic Press, New York 1988.

[15] L. Vincent, "Algorithmes morphologiques a base de files d'attente et de lacets , Extension aux graphes," *Ph.D. Thesis*, Paris School of Mines, May 1990.

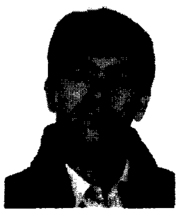
[16] F. Meyer, "Contrast feature extraction," *Quantitative Analysis of Microstructures in Materials Science, Biology and Medicine*, pp. 795–812, Riederer Verlag, 1977.

[17] F. Meyer, "Morphological image segmentation for coding," *Int. Workshop on Mathematical Morphology and its Applications to Signal Processing*, pp. 46–51. Barcelona, May 1993.

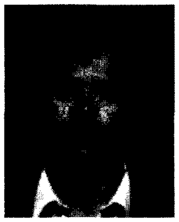
[18] Y. Yasuda, "Overview of digital facsimile coding techniques in Japan," *Proc. IEEE*, vol. 68, no. 7, pp. 830–845, Jul. 1980.

[19] D. E. Pearson and J. A. Robinson, "Visual comunication at very low data rates," *Proc. of the IEEE*, vol. 73, no. 4, pp. 795–812, Apr. 1985.

[20] J. R. Casas, L. Torres and M. Jareno. "Efficient coding of residual images," *Proc. of SPIE Visual Comm. and Image Processing*. vol. SPIE2094, Boston, MA, 1993



Josep R. Casas graduated in telecommunication Engineering from the *Escola Tècnica Superior d'Enginyeria de Telecomunicació* of the *Universitat Politècnica de Catalunya* (UPC) in 1990. He is Assistant Professor of Television Systems and Image Transmission courses at UPC, where he is currently working on his Ph.D. thesis. His research interests are in image and video coding, morphological image processing, image segmentation and advanced television systems.



Luis Torres received the Telecommunication Engineer degree from the *Universitat Politècnica de Catalunya*, Barcelona, in 1977, and the Ph.D. from the Electrical Engineering Department of the University of Wyoming, USA, in 1986. He is Associate Professor of the Polytechnical University of Catalonia, where he teaches Television Systems and Image Processing courses. He is currently responsible for several projects in low bit rate video coding applications working toward the MPEG 4 standard. His main interests are image coding, statistical coding, multiresolution schemes and object based systems.