RESIDUAL IMAGE CODING
USING MATHEMATICAL MORPHOLOGY *

Josep R. Casas, Luis Torres

Dept. of Signal Theory and Communications
Universitat Politècnica de Catalunya
P.O. Box 30.002
08071 Barcelona, Spain

ABSTRACT

A new method for the extraction, selection and coding of perceptually significant image components in a video sequence is presented. It has been applied to the coding residues of a segmentation-based multilevel coding scheme in order to show the need for a careful perceptual selection in very low bit-rate video coding applications. Mathematical Morphology operators are used for the extraction and selection steps of the proposed algorithm. The locations of the selected image components are then efficiently coded with a positional coder using prediction in the temporal dimension.

Results show that for coding rates lower than 10 Kbit/s, perceptual selection becomes a critical issue in order to achieve a good coding rendition. At higher rates, the fact of considering perceptual factors also leads to higher compression for similar subjective image quality.

1. INTRODUCTION

In Progressive Image Coding, the cost in bits of the successive residual images being encoded does not always result in accordance with the subjective importance that such components may have. Usually, as long as further improvement in the reconstructed image quality is required, the coding becomes more inefficient. Beyond a certain reconstruction level, slight improvements demand significant amounts of information support (channel or storage capacity) because irrelevant information is encoded along with meaningful details. This problem arises in most coding methods providing scalability, from linear pyramids to multilevel second generation schemes.

We propose an extension to video sequence coding of the morphological algorithm for detail extraction presented in [1], which will be applied to the coding residue of a segmentation-based video coding scheme. The

---

extracted details will be ranked and selected according to perceptual parameters. A suitable technique derived from an extension of 2D-runlength coding to the temporal dimension is then used for coding the spatial-temporal details. For a segmentation-based progressive scheme [2] we will compare the results of "detail-coding" and pure segmentation-based coding.

This paper is organized as follows. The efficiency problem introduced above will be presented and illustrated in section 2. Sections 3 and 4 will be devoted to the morphological detail extraction and the perceptual ranking and selection. In section 5, the coding technique is briefly described and, finally, in section 6 some results are given for very-low bit-rate video coding before the conclusions.

2. EFFICIENCY IN THE CODING
OF RESIDUAL IMAGES

Doubling the bit-rate does not usually correspond with having twice the subjective information content in the coded images. Although subjective measurements are quite difficult to assess, it can be noticed that the resulting image quality is sometimes far away from this "perceptual doubling". Compare the two levels of coded images shown in fig. 1 as an example for still image coding, and the sample frame coded at two different levels of the method proposed in [2] for video coding in fig. 2. For the JPEG scheme, it appears that the visual rendition of the coding system has been higher for the information coded in the first image (fig. 1b at 0.5 bpp) than for the information added in the second image (fig. 1c at 0.5+0.5=1 bpp). In the segmentation-based coding scheme this performance is also encountered. Here, the contour-texture coding technique becomes inefficient because non-significant regions are encoded while some perceptually significant regions have not been obtained in the segmentation process for fig. 2c.

Therefore, the selection of the information to be coded is a key issue in the coding scheme. It may be useful to have the image components represented separately, as regions are in a contour-texture scheme, in order to select only the most significant ones, and the selection criteria should be as close as possible to their visual perception.

Figure 1: a) original (left), b) JPEG coded at 0.5 bpp (center), c) JPEG coded at 1 bpp (right)

Figure 2: a) sample frame from the original sequence in QCIF format (left), b) segmentation-based coding [2] at 6 Kbit/s (center), c) segmentation-based coding [2] at 11 Kbit/s (right)

3. DETAIL EXTRACTION FROM RESIDUAL IMAGES

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 [3] provides tools that give a good insight into the structure of the image for processing purposes.

The morphological top-hat transform is computed as the residue or difference between the identity operator and the morphological opening of size 'n': $I-\gamma_n$'. It extracts bright contrasted components from the original image that are 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. If the filtering, i.e. morphological opening, is followed by a reconstruction process [3], the contours of the remaining objects in the filtered image are preserved

so that the result of the new reconstruction top-hat, 'I–$\gamma^{(rec)}$', does not contain spurious components due to edge variations in the bigger objects. Nevertheless, the grey levels of the extracted details are dimmer than what they should be, because part of their amplitude remains in the filtered image.

To overcome these drawbacks, a new morphological transformation was proposed [1], based on the selection of the significant bright contrasted details from the reconstruction top-hat and the computation of their true amplitude values from the morphological top-hat, by the geodesic reconstruction of a marker image that indicates the location of these desired components under the morphological top-hat.

We have applied the morphological detail extraction technique in the three dimensional space given by the two spatial and the temporal axis. Moreover, it has been slightly modified in order to obtain the details in a more suitable fashion for the posterior coding in a multilevel segmentation framework for video sequences [2]. In this progressive segmentation scheme, the information to be coded in the next level consists of new regions that are extracted from the current "coding residue". The coding

residue being defined as the difference between the original image and the reconstructed image at the previous coding level of the progressive hierarchy. The detail extraction algorithm should be able to identify non-extracted details from the current coding residue. Therefore, the markers for significant details are obtained from the coding residue instead of the original image.

If 'f' is the original image or video sequence and 'cod' stands for the current coded image, then the coding residue 'res = f − cod', is used in order to get a marker image 'mrk$_W$' for the bright details (dark details are obtained using dual operators, i.e., closing). A contrast threshold 'kσ' is applied to the residual image in order to select enough contrasted components. A second contrast threshold 'σ', smaller than the previous one, is used to avoid noisy low-contrast blobs connected with selected details, so that the reconstruction process takes place only under the points of the morphological top-hat that are contrasted enough, and does not propagate into connected spurious blobs. Both thresholds are made adaptive to the images being processed by setting their values to some multiple of the variance of the residue.

- morph. top-hat (original) and rec. top-hat (residue):

$$\text{tht}(f) = f - \gamma_n(f) \qquad (1)$$

$$\text{tht}^{(rec)}(\text{res}) = \text{res} - \gamma^{(rec)}(\varepsilon_n(\text{res}), \text{res}) \qquad (2)$$

- marker for bright details, mask for the background:

$$\text{mrk}_W = \begin{cases} 255 & \text{if tht}^{(rec)}(\text{res}) > k\sigma \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

$$\text{msk}_W = \begin{cases} \text{tht}(f) & \text{if tht}^{(rec)}(\text{res}) > \sigma \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

- image of bright details and smoothed image:

$$\text{det}_W = \gamma^{(rec)}(\text{mrk}_W, \text{msk}_W) \qquad (5)$$

$$\text{sm}_W = f - \text{det}_W \qquad (6)$$

The dual morphological transforms are applied on the smoothed image 'sm$_W$', in order to obtain the image of dark details 'det$_b$'. This sequence of operations yields a decomposition in detail image 'det = det$_W$ − det$_b$' and smoothed image 'sm = sm$_W$ + det$_b$' parameterized by size and contrast; the size of the structuring element, 'n', and the thresholds 'σ' and 'kσ', being the control parameters.

4. DETAIL RANKING AND SELECTION

The isolated components of the detail image are called details. The criteria used for the selection of these details 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. Size is an intrinsic parameter, whereas contrast is an extrinsic one since it is related to the background where the detail is placed. Besides contrast, some other extrinsic parameters will be considered. According to Weber's law, for the same contrast level details are more visible in dark areas than in bright 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. A third extrinsic parameter will thus be considered: the activity of the surrounding area.

The contrast parameter is the average amplitude of the pixels for each component of the detail image (a 'detail'). The grey level of the local background is the average amplitude over the set of pixels on the smoothed image underneath each detail. The background activity measure may be computed as the average of the morphological gradient [3] in the close neighborhood of each detail, i.e., in the area within some pixels distance 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 level and the activity measure, the ranking order is computed by the empirical formula:

- Ranking parameter: $\text{rnk} = \dfrac{\text{cn}}{\alpha \cdot \text{b}_L + \beta \cdot \text{b}_A} \qquad (7)$

Further terms may be included in eq. 7, 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. 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 details of the ranked list are selected for coding, depending on the available bit-rate.

5. CODING THE SELECTED FEATURES

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 PCM coded with eight bits and stored in a buffer of grey levels in scanning order. The shape and location of these details are coded as a binary image of detail supports using multidimensional run-length coding [4]. Differential runs are then obtained for the relative addressing of each detail and for relative spatial and temporal variations of its shape. In the video sequence, the reference line used for the coding of the current line is taken from the previous frame, so that the temporal correlation of image details is exploited by taking the runs in the temporal direction −3D run-length coding.

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.

Figure 3: a) sample frame from an original sequence (left), b) segmentation-based coding [2] using 160 regions (center), c) segmentation-based coding plus 40 volumic details (right)
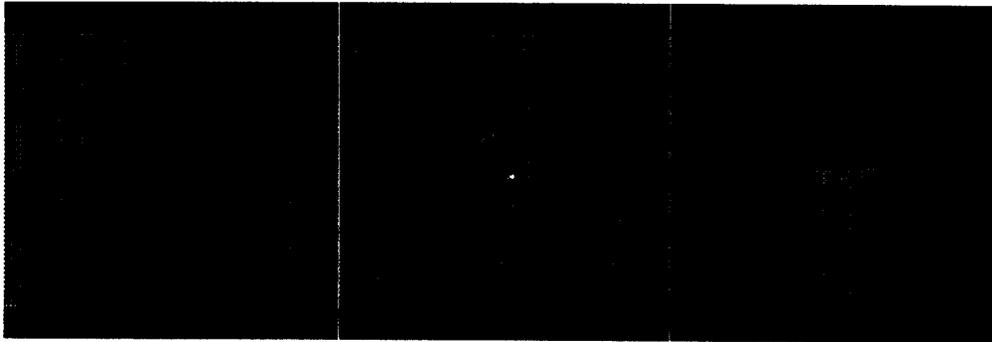

Figure 4: a) coding residue from figure 2b (left), b) 18 selected details (center), c) reconstruction at 10 Kbit/s (right)

6. RESULTS AND CONCLUSION

The proposed technique may be applied to still images but we present some results for video sequences as a more general case. Fig. 3 illustrates the 3D-detail extraction technique. Fig. 3b shows a sample frame from the segmentation [2] with 162 volumic regions in 10 frames. In fig. 3c, some details extracted from the coding residue have been added to the segmented sequence. Note the perceptual significance of the extracted details in the eyes, the nose, the phone or the letters in the wall.

In fig. 4, the detail coding technique has been applied to the lower coding level presented in fig. 2b. From the coding residue (fig.4a) some meaningful features have been obtained (fig.4b). The addition of these details to fig. 2b yields a new coding level (fig.4c) with a bit-rate similar to the higher segmentation level (fig.2c). It is worthwhile to observe the visual quality improvement that for a similar bit-rate represents the inclusion of the proposed technique. It has also been applied to other video sequences with the same results, what proves the goodness of the proposed algorithm.

We can conclude that the significance of some image features from a subjective point of view must be taken into account in second-generation coding techniques. This significance is even greater in progressive coding or for very low bit-rate video coding schemes. Mathematical Morphology provides suitable operators for the extraction of small video features, and an efficient algorithm has been described to this end. For the selection of the extracted details, a perceptually matched ranking method has been presented. Finally, the selected details are coded by means of a purpose-designed coding method.

References

[1] F.Meyer. *Morphological image segmentation for coding.* Workshop on Math. Morph. and its Applications to Signal Proc., pp. 46-51. Barcelona, May 1993.
[2] P.Salembier et al. *Morphological segmentation-based coding of image sequences.* IEEE Europ.Conf.on Circuits Theory and Design, Davos, Switzerland, Sept. 1993
[3] J.Serra. *Image analysis and mathematical morphology* Academic Press, New York 1982 (Vol1) and 1988 (Vol2)
[4] Y.Yasuda.*Overview of digital facsimile coding techniques in Japan.* Proc. IEEE, Vol.68,p830-845,1980.