OPEN ACCESS

University of BRISTOL

Link to published version (if available):
10.1109/ISCAS.1996.541803

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research

General rights

MORPHOLOGICAL SEGMENTATION BASED VIDEO CODING EMPLOYING CONDITIONAL SMOOTHING

P.J. Czerepinski and *D.R. Bull*
Image Communications Group, Centre for Communications Research,
University of Bristol, Queens Building, University Walk, Bristol BS8 1TR, UK
email: dave.bull@bristol.ac.uk p.j.czerepinski@bristol.ac.uk
Tel. +44 117 9288613; Fax: +44 117 9255265

Abstract In the context of coding the displaced frame difference signal, morphological segmentation is an alternative to linear methods such as the DCT. Although offering attractive features: edge preservation and lack of ringing, it can introduce artefacts such as spurious edges in the reconstructed video sequence. In this paper, a smoothing algorithm is described that reduces this problem and yields significant subjective quality improvement.

1 INTRODUCTION

The feasibility of compressing still images and video by exploiting the good correlation properties of image data and imperfections in the human visual system has been recognised for many years. Most existing standards for video compression employ the discrete cosine transform (DCT) for both intra- and interframe processing. It has been proved, however [1, 2], that the DCT is suited for coding highly correlated data only. In the interframe mode, motion compensation is usually performed prior to DCT coding. The correlation properties of the resulting displaced frame difference signal are relatively poor. High-frequency regions are produced around moving edges and these, if coded using linear transforms, cause ringing artefacts due to Gibbs phenomenon.

An alternative coding scheme, based on performing a morphological segmentation on the displaced frame difference signal has been proposed by Li and Kunt [3]. Morphological segmentation is a non-linear algorithm, which enables selecting or discarding frame-to-frame changes based on a shape/size criterion. Advantages of this method include a lack of ringing artefacts and edge preservation. It does however suffer from the fact that spurious edges can appear in the reconstructed image sequence corresponding to contours associated with the regions chosen for transmission. Although these are generally tolerable in high-contrast areas, they can be objectionable in regions of low contrast variation. In this paper a smoothing algorithm is introduced that mitigates this shortcoming and improves the subjective (and objective) performance of the morphological segmentation based coder.

2 MORPHOLOGICAL ALGORITHM FOR FRAME DIFFERENCE SEGMENTATION

2.1 Basic Morphological Operations

Elementary morphological operations are introduced here [4]. Let A and B be subsets of the $\Re^2$ plane; call A a binary image and B a binary structuring element. Then the dilation of A by B is defined as:

$$A \oplus B = \bigcup_{b \in B} (A + b) \qquad (1)$$

and the erosion of A by B is:

$$A \ominus B = \bigcap_{b \in B} (A - b) \qquad (2)$$

The geodesic reconstruction of A with M is given by:

$$\mathrm{rec}(A, M) = \mathrm{D}_A^\infty(M) \qquad (3)$$

where M is any subset of the image to be reconstructed and $\mathrm{D}_A^1(M)$ is the geodesic dilation:

$$\mathrm{D}_A^1(M) = (M \oplus B) \cap A \qquad (4)$$

and D_A^n denotes iterating the operation n times.

2.2 The Segmentation Algorithm

In this section, a morphological segmentation algorithm for the coding of the displaced frame difference signal based on the work of Li and Kunt [3] is described.

Thresholding

Let D_1 denote the displaced frame difference (figure 1a). At this stage, a binary mask D_2 is created, according to:

$$D_2 = \begin{cases} 1 & \text{if } |D_1(x,y)| > T \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

Pixels with a value of 1 are referred to as ON pixels and those valued 0 as OFF pixels.

Smoothing and Eliminating Single Pixels

The ON areas in image D_2 are irregular in shape (figure 1b). At this stage, filtering is carried out in order to obtain a set of regions with well-defined boundaries. Moreover, the thresholded frame difference contains a number of isolated pixels that will be removed here. This is accomplished by the following operations:

628

$$D_3 = \text{med}\big(\text{med}(D_2)\big) \tag{6}$$

$$D_4 = \text{rec}(D_2, D_3) \tag{7}$$

$$D_5 = \text{med}\big(\text{med}(D_4)\big) \tag{8}$$

where med denotes median taken in the 8-connected pixel neighbourhood.

Removing Small Regions

This is the key stage of the algorithm. It involves removal of those ON regions whose dimensions are considered sufficiently small and whose elimination does not significantly affect reconstruction quality for a given compression ratio. Firstly, the image D_5 (figure 1c) is eroded n times :

$$D_6 = \big(\underbrace{\ldots (D_5 \ominus B) \ominus B \ldots}_{n \text{ times}} \big) \tag{9}$$

Note, that this operation will remove any ON region not wider than $2n$. Then, reconstruction takes place, ensuring that large regions are retained and their shapes preserved. Thus:

$$D_7 = \text{rec}(D_5, D_6) \tag{10}$$
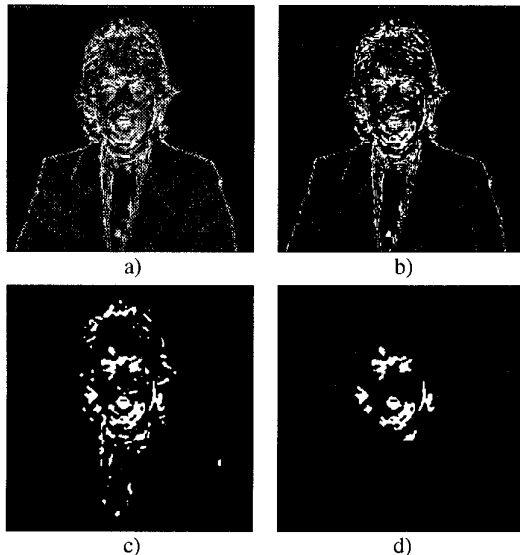

c)   d)

Figure 1 Stages of the segmentation algorithm. a) displaced frame difference, b) after thresholding, c) after smoothing and eliminating isolated pixels d) after eliminating small regions. Black denotes OFF and white ON regions.

Coding

Finally, the image D_8 is chosen for transmission according to:

$$D_8 = D_7 \circ D_1 \tag{11}$$

where $\circ$ represents scalar multiplication. D_8 is then subjected to quantisation and entropy coding.

Reconstruction quality and compression ratio can be controlled by varying the threshold value, T, the number of erosions, n, and the quantiser step size, Q.

3 THE SMOOTHING ALGORITHM

A disadvantage of morphological segmentation-based coding is due to the quantisation and thresholding processes. These result in step-wise colour intensity changes at the edges of transmitted regions, especially noticeable in areas of low colour intensity gradient. A simple solution to the problem, such as low-pass filtering the reconstructed frame, is not appropriate since it would inevitably remove some of the fine image detail. A better method involves conditional smoothing of the reconstructed frame [5], dependent on local variance criteria. This approach also has disadvantages, however: firstly, applying it to the whole frame, when only restricted areas are influenced, is inefficient from the implementation point of view. Secondly, the filtering may also affect areas which are never encoded, e.g. the still background. The method described here operates on the encoded areas only, keeping the computational cost low.

3.1 Statistical Basis of the Smoothing Operation

The method described here smoothes only the regions around the boundaries of the transmitted ON areas. The statistics of the OFF pixels in the D_8 image that have at least one ON neighbour have been investigated. Conditional distributions of the actual values of these pixels (which will henceforth be referred to as 'boundary pixels' - see figure 2), given a prediction $\hat{P}$ formed as the mean of the neighbouring ON pixel values, were evaluated.
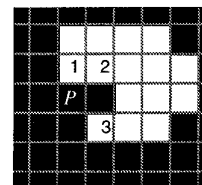

Figure 2 The value of the boundary pixel, P, is predicted from the values of the neighbouring ON pixels: 1, 2, 3.

Data was collected over 32 frames of the well known 'Claire' and 'Trevor' sequences. No quantisation was carried out on the transmitted signal in this case. From the graph in figure 3 it can be observed that, as long as the absolute value of the prediction, $|\hat{P}|$, is lower than the threshold value, $T=8$, the mean boundary pixel value is equal to half of the prediction. This is to be expected: if no statistical inquiry had been carried out, half the prediction value would have been the result of linear interpolation. Note however that, as $\hat{P}$ increases beyond the threshold, the mean boundary pixel intensity, $\overline{P}$, reaches a saturation

629

level, equal to $T/2$. This can be intuitively explained as follows: With the increase of $\hat{P}$ the increase of the boundary value P is expected. However, the larger the latter, the more likely it is to fall above the threshold value and be included into the ON area itself. This leads to the observed saturation effect. This has been confirmed for a variety of threshold values and, most importantly, for the case of quantised data, where a prediction is formed as the median of the neighbouring ON pixel values.
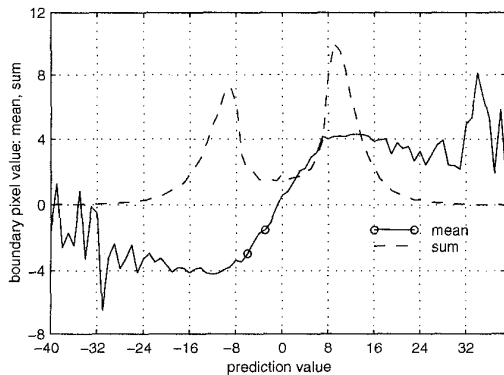
Figure 3 Mean boundary pixel value and normalised sum of boundary pixels in respect to prediction value. No quantisation was carried out. Prediction was formed as the mean of the touching ON pixel values. $T=8$.

As can be verified from the 'normalised sum' curve in figure 3, the distribution at high prediction values is derived from a vanishingly small number of samples, making the statistical analysis unreliable. However, high prediction values will coincide with high-contrast areas (e.g. moving edges), which should not be exposed to smoothing. This has lead to the development of the localised smoothing algorithm described below.

3.2 Implementation

A smoothing algorithm that takes advantage of the statistics presented above has been implemented. For every boundary pixel, the following substitution is made:

$$\tilde{P} = \begin{cases} \frac{1}{2}\hat{P} & \text{if } |\hat{P}| < T \\ \frac{1}{2}T\,\text{sign}(\hat{P}) & \text{if } T \leq |\hat{P}| < P_{max} \\ 0 & \text{if } P_{max} \leq |\hat{P}| \end{cases} \qquad (12)$$

By not modifying the boundary pixel values if $\hat{P}$ exceeds some arbitrary value, P_{max}, the algorithm ensures that naturally occurring high-contrast areas are not smoothed out.

4 CODING

The coding algorithm used consists of an intra-frame DCT or wavelet mode and a morphological segmentation based inter-frame mode. Both the encoder and the decoder (figure 4) perform localised smoothing around the transmitted data regions. In contrast to a simple post processing operation, this ensures that the benefits of the smoothing accumulate, rather than remaining local to individual frames. Block-matching motion estimation was used with the block size equal to 16 and half-pixel accuracy. In all experiments, the first frame was encoded at an identical signal to noise ratio of 35.55 dB. The differential signal D_8 was subjected to a uniform quantiser, the threshold value T was varied between 7 to 17 and P_{max} was set to 40. Motion vectors were encoded using a conditional model, depending on the median prediction formed from the three previously transmitted neighbours. A conditional model was also used in the case of the differential data D_8, depending on the state S of the encoder:

$$S = \begin{cases} S1 & \text{if } \forall V \in \{X,\ Y,\ Z\},\ V = \text{ON} \\ S2 & \text{if } \exists V \in \{X,\ Y,\ Z\},\ V = \text{OFF} \end{cases} \qquad (13)$$

where X, Y, Z are the causal half-plane neighbours of pixel at location V. In both cases, fixed statistical models were used.
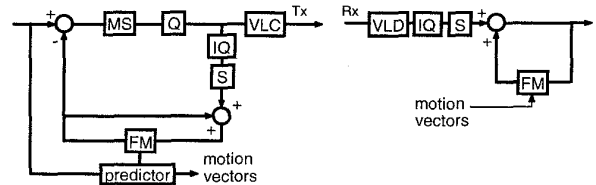
Figure 4 Codec block diagram (inter-frame mode). MS - morphological segmentation; Q - quantiser; IQ - inverse quantiser; S - smoothing; FM - frame memory; VLC - variable length coding; VLD - variable length decoding.

5 RESULTS

The morphological segmentation algorithm was applied to the initial 32 frames of the Trevor sequence, firstly in its original form and secondly using the smoothing algorithm. As can be verified from the graph in figure 5, the smoothed version gave improvements in both signal to noise ratio and compression ratio for fixed segmentation parameters. If methods are compared at identical compression ratios, it can be seen that the smoothing operation yields an improvement of approximately 0.4 dB. Examples of subjective results are depicted in figure 6. The patches formed by the updated areas in the case of no smoothing are clearly visible, as is the improvement in quality achieved from applying the smoothing algorithm.
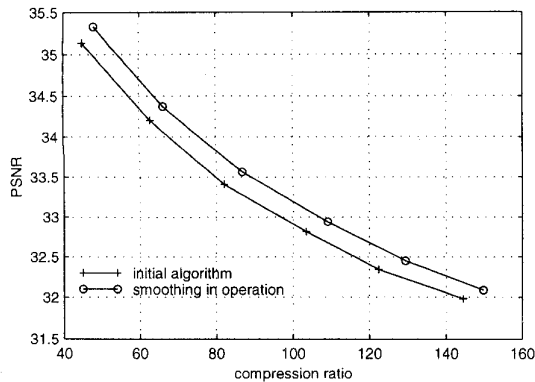
630

Figure 5 Coder performance for the Trevor sequence
(256×256×8 bpp), interframes only.

6   CONCLUSIONS

A smoothing technique is proposed, that improves the performance of a morphological segmentation based video coder. Being based on a local prediction, the algorithm is capable of modifying the reconstructed image sequence using data which is not explicitly available in the transmitted bit-stream. At identical bit-rates, an 0.4 dB improvement is reported over the initial algorithm. The only overhead required is associated with the transmission of the threshold value, T.

REFERENCES

[1] Flickner M.D., Ahmed N. *A Derivation for the Discrete Cosine Transform*, Proc. IEEE, v.70, no.9, Sept. 1982, pp.1132-34.

[2] Jain A.K. *A Sinusoidal Family of Unitary Transforms*, IEEE Trans. PAMI, v.1, Oct. 1979, pp.356-365.

[3] Li Wei, Kunt M. *Morphological Segmentation Applied to Displaced Frame Difference Coding*, Signal Processing v.38, 1994, pp. 45-56.

[4] Giardina C.R., Dougherty E.R. *Morphological Methods in Image and Signal Processing*, Prentice Hall, New Jersey 1988.

[5] Li Wei, Bhaskaran V. *A Low Bit Rate Video Codec*, HP Laboratory Report HPL-93102, November 1993.
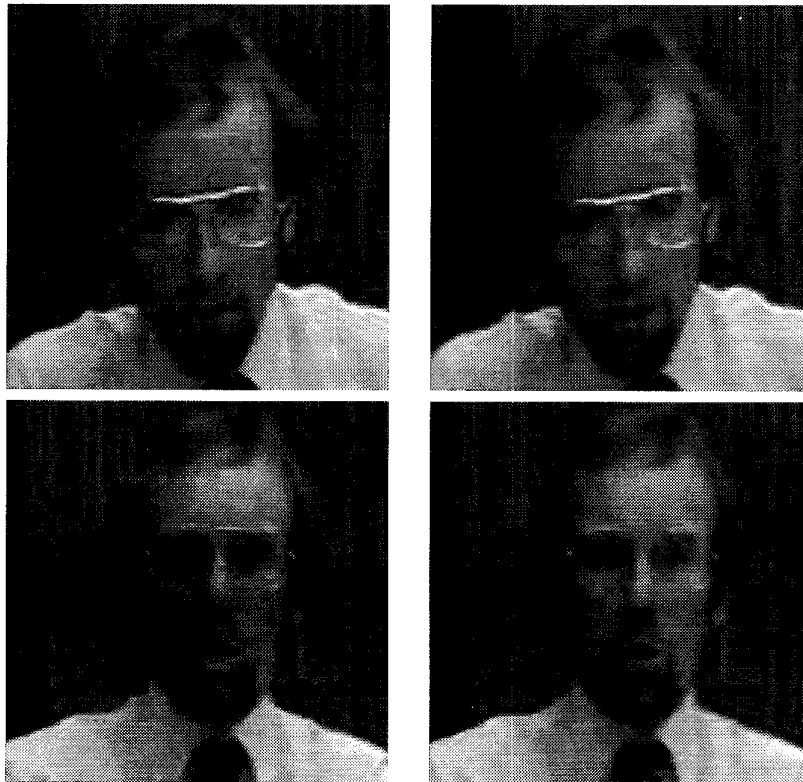
Figure 6 Subjective results; 128×128 zoom into 'Trevor' at a compression ratio of 63:1 , top left to bottom right: initial algorithm (frame 5), using smoothing (frame 5), initial algorithm (frame 25), using smoothing (frame 25).

631