CODING OF DIGITAL VIDEO WITH THE EDGE-SENSITIVE DISCRETE WAVELET TRANSFORM

Kalman Cinkler

University of Bremen, Telecomm. Dept. (FB1)
Kufsteiner Str. NW1, D-28359 Bremen, Germany
http://www.comm.uni-bremen.de/whomes/cinkler
cinkler@comm.uni-bremen.de

Alfred Mertins

Technical Faculty of the
Christian-Albrechts-University in Kiel, Germany
Telecommunications Institute
mertins@techfak.uni-kiel.de

ABSTRACT

In this paper we present an image sequence coding scheme for very low bit rate coding which is based on spatial redundancy reduction via the new edge sensitive subband coding method and temporal redundancy reduction via windowed overlapped block-matching motion compensation. The scheme has the main advantage, that only the significant regions of difference-images are coded. Thus the computational cost can be kept low. Due to the properties of both the temporal and the spatial coding used, there are no blocking effects in the coded images, and the overall visual performance of the coding scheme is very good.

1. INTRODUCTION

The requirements on video source coding algorithms has enormously increased in the past years due to many factors, one of them being the availability of hardware capable of processing vast amounts of visual data in real-time. Some image sequence coding schemes for various bit rates have already been internationally standardized: MPEG-2 for television broadcasting at 10 Mb/s, MPEG-1 for digital video at 2 Mb/s, H.261 for video-conferencing at n×64 kb/s and H.263 for video-conferencing at very low bit rates, i.e. below 20 kb/s.

Since the three-dimensional transforms are proven not to be quite efficient, the coding of an image sequence is usually performed in two steps: temporal redundancy reduction and spatial redundancy reduction. The temporal–spatial order from above is the most often applied one in very low bit rate coding. In typical applications such as video-telephony the frame-to-frame change is relatively slow (in some areas of the image even absent, e.g. background) and the temporal redundancy reduction can be performed with great efficacy. After this, only the differences between predicted and current frames remain. These difference-images are coded with an adequate two-dimensional transform in order to remove the spatial redundancy.

All the standards for coding of digital video mentioned above employ the discrete cosine transform (DCT) as the transform for spatial coding. The DCT itself, in the way it is implemented in these standards is a rather inefficient coding scheme in terms of coding gain. The blocks which the image is broken into prior to processing (e.g. [1]) have to be small in order to be represented with reasonable accuracy. However these blocks are processed independently from each other, and the inter-block redundancy remains unutilized. Such a block-wise processing has an unimpeachable advantage though: it inherently matches the block-nature of the most popular motion compensation methods, i.e. the block-matching motion compensation methods.

When using the discrete wavelet transform (DWT) instead of the DCT in order to reach even lower bit rates, one faces two problems. After the block-matching motion compensation, the difference-images contain sharp edges on block boundaries. Such edges in the input-image for the DWT decrease the achievable coding gain. The second problem follows from the need for spatial selectivity of the picture-refreshing strategy during the encoding. For a given (very low) bit rate it is not possible to code every difference-image entirely. According to some appropriate strategy, just significant areas of difference-images (see Fig. 1c) are coded in our method. Note that the DWT cannot be applied to this problem in its original form.

In the following sections these problems are addressed separately, followed by the description of the overall coding scheme and by results of coding.

2. TEMPORAL CODING

In conventional block-matching motion compensation (BMMC) schemes the picture is segmented into distinct (non-overlapping) blocks, and one motion vector

(a) Motion-vector field (b) Edge-map (c) Significant regions

Figure 1: The significance-decision strategy.

per block is obtained using block-matching methods. The prediction of the current picture is obtained from the previous picture and the motion vectors. Apart from the trivial case of identical motion vectors over the entire picture, this method produces a predicted image with visible block boundaries. Therefore, the difference between the predicted and the current picture also contains sharp edges. When applying the DWT to the difference-image, the sharp block-edges cause large transform-coefficients in the higher frequency bands, and the achievable coding gain decreases.

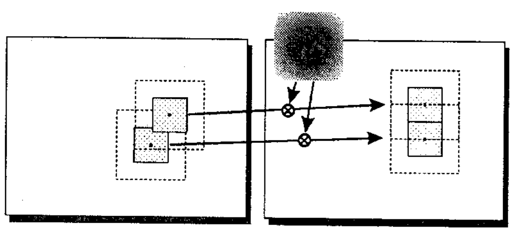

Figure 2: The windowed overlapping block-matching motion compensation.

For the video coding scheme presented in this paper, we used the well known overlapping BMMC with block-windowing as shown in Fig. 2, see [2] and [3]. The blocks for which the motion vectors are obtained are larger, they overlap, and prior to merging them into the predicted image, they are windowed by an appropriate window function. In this way, no edges are present in the predicted image. The difference image is smooth too, allowing high compression ratios. An example of a motion-vector field is shown in Fig. 1a.

3. SPATIAL CODING

As already pointed out in the introduction, it is not possible to code arbitrarily shaped regions of a difference image (e.g. in Fig. 1c) with the DWT in its original form. In [4], we presented a very efficient edge preserving image coding technique based on the DWT, the edge sensitive subband coding (ESSBC), which we apply here.

The ESSBC utilizes the edge map of the processed image (which not necessarily contains closed curves only, see Fig. 1b) in the way, that no filtering over these edges is performed.

Edge-detection is a simple task, it can be performed as a one-pass 2D-filtering. After the edges have been detected (e.g. as in Fig. 1b), the filtering for the subband decomposition is performed separately for the rows and columns of the image. This filtering is interrupted on edges. This means, if a row of the image is split by k edges into $k+1$ sections, each of these sections is filtered and downsampled separately. In order to perform support preservative decompositions of arbitrary-length sections we use linear phase filters that allow symmetric reflection at the section boundaries. Depending on the length of the sections (even or odd), and whether they start at even or odd positions, there are four cases of processing (see [4] for implementation details). Processing all rows of the image in the described way (section-by-section), we get the same number of coefficients in the two subbands regardless of the number and length of the sections. This is very important for the next decomposition which is performed on the columns of the subbands. Such a decomposition results in four subbands and we repeated it on the respective lowpass-lowpass subbands twice in order to get ten subbands altogether. To provide the later decompositions with the edge-position information, the edge map has to be downsampled correspondingly.

Due to the fact that no filtering over sharp edges is performed, the power in the higher frequency bands is considerably lower than for the Standard DWT. This enables a very high compression ratio, since much more coefficients in the higher subbands are quantized to zero than in the case of the Standard DWT. To perform the synthesis in the receiver, we need the transform-coefficients and the edge map as well. The edge map can be compressed very efficiently though, and despite this additional information the achievable overall compression ratio is still higher than for the Standard DWT at the same level of visual quality.

Once we are able to decompose images with any edges, we can think of the boundaries of the difference image's significant regions (Fig. 1c) as of additional edges. Since the boundaries of these regions build closed curves (closed "edges") and since no filtering over edges is performed, the regions are independent of the rest of the image, which gives us the freedom of simply leaving out the insignificant areas of the image and keeping the significant ones only.

4. THE CODING SCHEME

The entire coding scheme is shown in Fig. 3. The estimation of the motion vectors is performed on original images only, so that vectors having high correspondence with the real motion in the scene are obtained. This correspondence is valuable when performing motion compensated interpolation in order to enhance the frame rate at the decoder side. The task of compensation of quantization effects and noise, which is usually partly performed by motion compensation, we entirely left to the spatial coding, i.e. to the ESSBC. This is a good decision in so far as the ESSBC can handle local changes much more precisely than the BMMC.

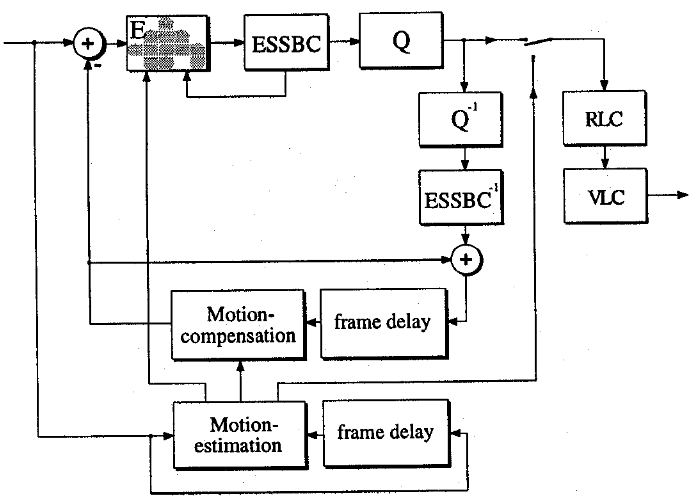

Figure 3: The coding scheme.

In Fig. 3, the block "Q^{-1}" stands for rescaling of coefficients, "$ESSBC^{-1}$" for the edge sensitive synthesis. "RLC" and "VLC" mean run-length coding and variable length (Huffman) coding respectively.

The block "E" represents the significance-decision, i.e. the decision about the region that will be spatially coded. This decision is made based on two information sources: the motion vector field and the edge map of the image. In areas where motion is present, we have to refresh more often than on background. The overall spectral characteristic of the scheme is a lowpass characteristic, which means that sharp edges in the scene can get blurred. Therefore, it is convenient to refresh parts of the image near these edges more often. We can think of the significance-decision as of merging the matrices in Figures 1a and 1b and making a threshold-decision for a given bit rate and a given quality (i.e. quantizer-step). The result of such a decision may be the region shown in Fig. 1c.

We performed comparative tests between our coding scheme and the emerging standard H.263 (a classical MC-DCT based coding scheme for very low bit rate coding, [5]) on various test sequences. The frame rate was 5 fps, we used image-sequences in QCI-format, and the target bit rate was chosen to be 10 kbps. Due to the edge preserving property of the edge sensitive subband coding, our coding scheme outperforms the H.263 in terms of visual quality. The coded images are sharper, more detailed and clear, without blocking effects, see Fig. 4.

5. CONCLUSION

The presented image sequence coder is a very efficient coder, it is especially suited for very low bit rate coding. It is simple, but powerful. It produces very good image quality at moderate computational costs which are comparable to the computational costs of conventional MC-DCT coding schemes. Due to the properties of the coding used, there are no blocking effects.

6. REFERENCES

[1] D. Le Gall: "MPEG: A Video Compression Standard for Multimedia Applications", Communications of the ACM, 34(4):46–58, April 1991.

[2] S.Nogaki and M.Ohta: "An overlapped block motion compensation for high quality motion picture coding", in Proc. of IEEE Int. Symp. Circuits and Systems ISCAS'92, Vol. 1, pp. 184-187, 1992.

[3] H.Watanabe and S.Sighal: "Windowed Motion Compensation", in Proc. of SPIE Visual Commun. Image Proc. Vol. 1605, pp. 582-589, Nov. 1991.

[4] K.Cinkler and A.Mertins: "Edge Sensitive Subband Coding", in Proc. of IEEE Int. Conf. Acoust. Speech and Sig. Proc. ICASSP'96, Atlanta, May 1996.

[5] http://www.fou.telenor.no/brukere/DVC/h263_software/

Figure 4: Comparison of the performance of the standard H.263 and our method on the sequence "Suzie": (a), (c) and (e): three frames (No. 45, 47 and 49 respectively) coded with H.263, and (b), (d) and (f): the same frames coded with our method. The bit rate is 10 kbps for QCIF images @ 5 fps.