Adaptive Wavelet-Packet Decomposition for Rate Control of Object Oriented Coding of Video Sequences

Ioannis Stephanakis, Anastasios Douiamis, Nikolaos Douiamis and Stefanos Kollias
National Technical University of Athens, Department of Electrical and Computer Engineering
Heroon Polyteknion 9, 157 73 Zografou, Greece
E-mail: {stephan,adoulam,ndoulam}@image.ece.ntua.gr, stefanos@cs.ntua.gr

Abstract

A novel object dependent coding scheme based on optimization upon wavelet packet trees is proposed. The method may be employed in the context of object oriented video standard MPEG-4 for constraint distortion minimization for each Video Object (VO). Optimal transmission rates are evaluated for each VO which guaranty that important VO (like the foreground) are transmitted with distortions which are equal or less than preselected distortion values. The proposed algorithm is applied to object dependent coding of video frames of the Akiyo sequence with Peak-Signal-to-Noise (PSNR) values which are higher for the foreground than the background. Experimental results which favor the proposed method are provided as well.

1. Introduction

The increasing demand for video transmission as well as the need of providing simultaneously multimedia capabilities to future encoders have created the standard called MPEG-4, [1], which permits high compression ratios, interactivity, universal accessibility and portability of video content. The new standard adopts the concept of Video Objects (VOs) and Video Object Planes (VOPs) which are simply the projection of a VO onto a frame. Any information related to each VO such as shape, location, motion, and texture is coded into a separate layer (Video Object Layer) and embedded into the bit stream. Therefore, separate decoding and composition of VOPs at different object rates and resolutions are permitted by directly examining the MPEG stream. Region-based coding using the block based DCT may be utilized to encode each VO using different quality factors. Different quality factors allow for enhancement of the perceptual quality of the reconstructed images according to the importance of the objects, e.g., the object which represents the foreground of the image is coded with a higher quality factor than the object which represents the background of the image [2]. Bit allocation techniques [3] may be adopted to determine the bit allocation among VO.

Wavelet packet transforms provide good spatial and frequency localization. This makes them suitable for image compression [4], [5] and image processing applications [6]. They are to be adopted into the coding standard for still frame images JPEG-2000 whereas several methods employing wavelet packet transforms for encoding video have been appeared recently in the literature [7], [8]. Wavelet packets defined upon orthogonal masks allow for different orthogonal transforms within the block of the image object. The coefficients of such transforms are computed iteratively and are very efficient in rendering the texture of the object.

Algorithms of optimal tree selection, like those in [3], [9],[10] which determine the best wavelet packet transform, may be used for object depended coding. Such an algorithm is proposed in this paper. The proposed method is equivalent to designing object depended masks which take into consideration the statistics of each object per Group of Pictures (GOP). Adaptive optimization is performed for each object and GOP so that total distortion per frame is minimized while the distortion of important objects like the foreground is kept below a preselected constraint. The optimal tree is efficiently transmitted for each object and GOP with only a few bits while versatility of the encoder compared with standard spatially adaptive DCT is increased dramatically. Results indicate that distortion-rate PSNR curves are better than the corresponding distortion-rate curves using the DCT by 3dB or more. A variable length entropy coding scheme is adopted but extension of the proposed technique by adopting Embedded-Zero-Tree (EZT), [11], coding is possible.

2. System Architecture for Perceptual Object Oriented Wavelet Video Coding

Figure 1 illustrates the architecture of the proposed perceptual wavelet video coding. The image in the Figure 1 is first segmented into semantic video objects (VOs) by applying a video object segmentation algorithm. A different degree of distortion is imposed on each video object according to its significance. In particular, objects
of high importance are distorted as less as possible while objects of low importance are quantized with larger quantization steps. Thus, more bits are allocated to the important objects, i.e., the foreground regions of an image, where the human visual system is more sensitive than to the unimportant areas. This scheme is in agreement with the concepts of the new MPEG-4 standard where the video stream is partitioned into several video objects with different importance being imposed upon each of them.

An optimal wavelet packet decomposition is calculated for each video object so that more efficient coding is achieved. The optimal wavelet packet is estimated by pruning the wavelet decomposition tree so that nodes corresponding to higher distortion (less quality) are eliminated. This optimal decomposition actually depends on the frequency components of the respective object and the anticipated transmission rate. For instance, all nodes of the wavelet decomposition tree corresponding to high frequencies are eliminated in case of an object with low frequency components. On the other hand, nodes of low frequency are retained.

Both I-frames and P-frames are included in the proposed coding algorithm similarly to the conventional MPEG standards. In particular, I-frames are generated without any reference to other frames. On the contrary, P-frames are coded with reference to the previous I or P-frame. In this case, the wavelet packet decomposition is applied to the frame after motion compensation.

3. Object Based Rate Control

We assume that we have N objects \( \{O_1, O_2, \ldots, O_N\} \) which are transmitted by N GOP consisting of an intraframe (I) at instance \( t \) and several predicted (P) frames at instances \( t+1, t+2, t+3, \ldots \). We further assume that the first \( M < N \) objects have to be quantized and coded in such a way that the total distortion \( D(Q_{m}(t=I), Q_{m}(t+1=P), Q_{m}(t+2=P), \ldots) \) for the GOP corresponding to \( O_m \) in an advanced MPEG system at instances \( t, t+1, t+2, \ldots \) is less or equal to \( D_m \), \( m \in \{1, 2, \ldots, M\} \). One quantizer, denoted as \( Q_m \), is used to quantize object \( O_m \) throughout the entire GOP. Our goal is to minimizing the total distortion for all GOP corresponding to video objects \( O_m, m \in \{1, 2, \ldots, N\} \) in such a way that a rate constraint over all objects is satisfied. Thus, the constraint optimization problem reads

\[
D_{T,GOP} = \min_{\text{all quantizers}} \left\{ \sum_{m=1,2,\ldots,N} (Q_{m}(t=I), Q_{m}(t+1=P)) \right\}
\]

subject to

\[
D_{m,GOP} = D(Q_{m}(t=I), Q_{m}(t+1=P), \ldots) \leq D_m \] and

\[
R_{T,GOP} = \sum_{m=1,2,\ldots,M} R_{m,GOP} \leq R_T, \text{ where } m \in \{1, 2, \ldots, M\}
\]

One may rearrange the first \( M \) video objects which are transmitted with low distortion over the GOP in such a way that \( \lambda_1(D_1) \leq \lambda_2(D_2) \leq \ldots \leq \lambda_M(D_M) \) with \( \lambda_m(D_m), m \in \{1, 2, \ldots, M\} \), being the negative of the slope of the distortion-rate curve corresponding to object \( O_m \) for a GOP at distortion \( D_m \). Typical objects in a video sequence are the foreground and the background. Different wavelet trees correspond to different transmission rates. For such an arrangement of \( \lambda_m \)'s, the solution of the above constraint optimization problem at
the convex hull may be found by the following optimization in which a Lagrange multiplier \( \lambda \) is used,

\[
J = \min_{\text{all quantizers}} \left\{ \sum_{n=1}^{M} D_n(t+1), Q_n(t+1+P) \ldots + \lambda \sum_{n=1}^{M} R_n(t+1), Q_n(t+1+P) \ldots \right\}
\]

(2)

where \( M' \leq M < N \) subject to \( \lambda_1(D_1) \leq \lambda_2(D_2) \leq \ldots \leq \lambda_{M'}(D_{M'}) \leq \lambda \leq \lambda_{M+1}(D_{M+1}) \ldots \leq \lambda_{M}(D_{M}) \) and \( R_T \leq R_T' \). The rate corresponding to slope \( -\lambda \) in the distortion-rate curve of object \( Q_t \) for a GOP is denoted by \( R_{n,\text{GOP}}(\lambda \alpha) \). The Lagrange multiplier \( \lambda \) is used to take care the rate constraint.

It is possible to solve iteratively the optimization problem presented in Eq. (1) on the overall convex hull by finding the \( (R_{n,\text{GOP}}(\lambda), D_{n,\text{GOP}}(\lambda)) \) pairs which satisfy Eq. (2) as follows:

**Step 1:**
For \( m \in [1,2,\ldots,M] \) find \( \lambda_m(D_m) \) and \( R_{n,\text{GOP}}(\lambda_m) \). Set \( R_T' = R_T - \sum_{n=M}^{n=1} R_{n,\text{GOP}}(\lambda_m) \). If \( R_T' \) is negative the maximum allowed rate is exceeded and there is no solution of the optimization problem. Otherwise set \( M' = M \) and continue.

**Step 2:**
Solve the minimization problem

\[
J = \min_{\lambda} \sum_{M' \leq l \leq N} \left( D_{l,\text{GOP}} + \lambda R_{l,\text{GOP}} \right)
\]

subject to \( \sum_{n=M' \leq l \leq N} R_{l,\text{GOP}}(\lambda) \leq R_T' \) for current \( M' \) and \( R_T' \). Find new \( \lambda \).

**Step 3:**
If \( \lambda_1(D_1) \leq \lambda_2(D_2) \leq \ldots \leq \lambda_{M'}(D_{M'}) \leq \lambda \leq \lambda_{M+1}(D_{M+1}) \leq \ldots \leq \lambda_{M}(D_{M}) \) then stop, otherwise set \( M' \) equal to smallest \( m \) for which \( \lambda \leq \lambda_m(D_m) \), \( m \in [1,2,\ldots,M] \), and set \( R_T = R_T - \sum_{n=M' \leq m} R_{n,\text{GOP}}(\lambda_m) \). Go to Step 2.

4. **Experimental Results**

The sequence of ‘Akiyo’ images with dimensions 144x176 pixels is used to perform numerical experiments using both I- and P- frames. Uniform frame coding using the standard DCT and the optimal WPT with various mask sizes (8x8, 16x16, 32x32, 64x64) is carried out first for an I-frame of Akiyo. This original frame is depicted in Figure 8(a). The PSNR relationships for the WPT and DCT are presented in Figure 2 for different values of bit rate. The larger the mask size, the higher the PSNR value. Variable length coding is used to encode the WP coefficients of image blocks. The set of quantizers \( Q_n \) for video object \( O_n \) is defined upon all possible wavelet packet bases derived by a wavelet family (Daubechies of order one). There is a one-to-one correspondence between possible wavelet packet decompositions and trees \( S_n \). Each channel of the wavelet packet transform is separately quantized and \( Q_n(S_n) = \{ q_n(t_1), q_n(t_2), q_n(t_3) \ldots q_n(t_k) \} \).
where \( \{i_1, i_2, \ldots, i_L\} \in \mathcal{S} \) is the complete set of leaves or terminal nodes of \( S_m \). The quantization steps are \( \{Q_{n}(i_1), Q_{n}(i_2), \ldots, Q_{n}(i_L)\} \). This information may be considered as a generalization of the quality factor defined for DCT and should be transmitted along with the image. Thresholds in each channel of the transform equal the quantization steps.

The PSNR comparison between uniform and object-dependent coding (Object 1 is the foreground and Object 2 is the background) is presented in Figure 3 for I-frames and in Figure 4 for P-frames for the Akiyo sequence. It is observed that the conventional uniform DCT method provides the lowest PSNR for a given bit rate, while the proposed method provides the highest PSNR.
modified DCT-based rate control algorithm, proposed in one of our earlier works [2], is included in these figures for comparison. WPT provides better results than this technique which is depicted in Figures 3 and 4 as object dependent DCT. For P frames the mask size has been selected to be 8x8 since larger size provides no significant PSNR improvement.

The PSNR plots of the foreground and the background for P frames using WPT and DCT coding are presented in Figures 5 and 6. A significant increase is observed in all cases. We further reallocate bits from the background to foreground in order to improve coding efficiency. This results in an increase of the PSNR values of the foreground and a decrease of the PSNR values of the background. The overall PSNR plots are presented in Figure 7 (P frames). Initial OD in this figure stand for object dependent coding. However, beyond a limit, the improvement of the total quality deteriorates since the foreground improvement is smaller than the respective degradation of the background. Examining several values, we concluded that a 5dB PSNR increase of the foreground results in the best total performance.

![Optimal wavelet decomposition trees for P frames. (a) Uniform WPT. (b) Object dependent scheme for foreground. (c) Object dependent scheme for background.](image)

Coded I-frames and P-frames are presented in Figures 8-9 for uniform DCT, uniform WPT and object dependent coding. We assumed that the PSNR of the foreground is 5dB higher than the PSNR of the background in order to get the best increase of the overall PSNR curve. As we can see from these figures the proposed method (Figures 8(d) and 9(d)) provides much better image quality than the conventional approach (Figure 8(b) and 9(b)). Furthermore, coding results are better than the object dependent DCT approach presented in [2]. We observe that the proposed method improves the quality of both the foreground and the background regions since different wavelet decomposition trees are used for the two regions. Thus, coding is optimized for each object (region) not to the whole frame as in the conventional DCT approach. All rates are 0.15 bits/pixel. The different trees used in WPT in the case of P-frames are presented in Figure 10 (both uniform and object dependent).

5. Conclusion

Object dependent coding of video sequences based on variable length encoding of the coefficients of the optimal WPT outperforms standard DCT coding techniques. Information regarding the optimal tree may be transmitted efficiently according to the proposed scheme.

6. References