Improvements in DCT-based video coding

Puri, Atul, Schmidt, Robert, Haskell, Barry

Atul Puri, Robert L. Schmidt, Barry G. Haskell, "Improvements in DCT-based video coding," Proc. SPIE 3024, Visual Communications and Image Processing '97, (10 January 1997); doi: 10.1117/12.263279

Event: Electronic Imaging '97, 1997, San Jose, CA, United States

Improvements in DCT Based Video Coding

A. Puri, R. L. Schmidt and B. G. Haskell

AT&T Labs
101 Crawfords Corner Road
Holmdel, NJ 07733

ABSTRACT

We report on recent advances in traditional DCT based video coding at low bitrates. These improvements allow either an increase in coding efficiency or an increase in other functionalities. Our investigation is conducted within the framework of the ongoing work towards the MPEG-4 video standard. The ISO Moving Picture Experts Group (MPEG) is currently developing this standard after having completed the MPEG-1 and the MPEG-2 standards. The MPEG-4 video standard is addressing a number of content based as well as traditional functionalities. The development process consists of iterative refinement of the Verification Model (which describes the coding method) via a set of well defined core experiments.

Our first experiment is on improved coding efficiency of Intra and uses DC and AC predictions and optimized scanning of DCT coefficients followed by a separate optimized variable length code table. Our second experiment is the study of bidirectional coding to allow additional functionality such as temporal scalability at low bit-rates. We present results of these experiments and summarize our findings.

Keywords: video compression, video coding, MPEG coding standards, MPEG-4, interframe coding, DCT coding, motion compensation, motion compensated DCT coding.

1. INTRODUCTION

The recent experience in the development of the ITU-T H.263 standard illustrates that several incremental improvements when combined can collectively result in a notable advance in the coding performance or increase in functionality of an earlier coding standard (in this case, H.261) making it quite competetive to fairly complex new proposals. It appears that although the DCT based coding framework was optimized for H.263 for low bitrate coding, it can still be refined further to allow improved coding efficiency as well as increased functionality. In this paper our goal is to investigate techniques for further refinments within the context of the ongoing development work of the MPEG-4 video coding standard. The ISO Moving Picture Experts Group (MPEG) is currently developing this standard after having completed the MPEG-1 and the MPEG-2 standards.

The MPEG-1 video standard is designed for Digital Storage Media (DSM) applications at bitrates of about 1.2 Mbit/s and supports basic interactivity with stored bitstream such as random access, fast forward, fast reverse and others. MPEG-1 video coding [1] uses block motion compensated DCT coding within a group-of-pictures (GOP) structure consisting of an arrangement of intra (I-), predictive (P-) and bidirectional (B-) pictures to deliver good coding efficiency and desired interactivity. This standard is optimized for coding of noninterlaced video only. The second phase MPEG (MPEG-2) standard, on the other hand, is more generic. MPEG-2 is intended for coding of higher resolution video as compared to MPEG-1 and can deliver TV quality in range of 4 to 10 Mbit/s and HDTV quality in range of 15 to 30 Mbit/s. MPEG-2 video standard [2,4] is mainly optimized for coding of interlaced video. MPEG-2 video coding builds on the motion compensated DCT coding framework of MPEG-1 and further includes adaptations for efficient coding of interlaced video [3,5]. MPEG-2 supports interactivity functions of MPEG-1 as well as new functions such as scalability [3]. Scalability is the property that enables decodability of subsets of entire bitstream on decoders of less than full complexity to produce useful video from the same bitstream. Scalability in picture quality is supported via SNR scalability [3], scalability in spatial resolution by Spatial scalability [3] and scalability in temporal resolution via Temporal scalability [3,6]. The MPEG-2 video standard is both forward and backward compatible with MPEG-1; backward compatibility can be achieved using spatial scalability. Both MPEG-1 and MPEG-2 standards only specify a bitstream syntax and decoding semantics, allowing considerable innovation in optimization of encoding.

The MPEG-4 standard, was started in 1993 with the goal of very high compression coding at very low bitrates of 64 kbit/s or under. Coincidentally, the ITU-T also started two very low bit-rate video coding efforts: a short term effort to be completed by early 1996 intended to improve H.261 for coding at around 20 to 30 kbit/s, and a long term effort to be completed by late 1998 intended to achieve higher compression coding at similar bitrates. Currently, the ITU-T short term standard called H.263 [8] is complete and there is also an ongoing activity for improving this standard and the refined standard will be called the H.263+ standard. In the meantime, the ongoing MPEG-4 effort [9,10] is focussing on providing a new generation of interactivity with the audio-visual content, i.e., access to and manipulation of objects or in the coded representation of a scene. Furthermore, moderate improvement in the basic coding efficiency is also expected offset the overhead needed for content based and other functionalities. Up to now, optimization of MPEG-4 has been carried with noninterlaced video at bitrates in range of 10 kbit/s to 1000 kbit/s. However, MPEG-4 is also expected to address higher bitrates and other video formats. Among the foreseen applications of MPEG-4 are mobile video phone/game/information access terminal, video answering machines and video email, interactive multimedia over internet, video catalogs, home shopping, virtual travel, surveillance, networked games etc.

As mentioned earlier, the MPEG-4 video effort [9,10,14] is addressing a number of content based as well as traditional functionalities which can be grouped into three main areas - high compression, content based interactivity and universal accessibility. The development process consists of iterative refinement of the Verification Model (VM) which describes the coding method, via a set of well defined core experiments. The latest Verification Model is VM5.0 [15]. Recently, following MPEG-4 workplan, a first draft of the standard describing the formal bitstream syntax and the decoding process was also released and is referred to as WD1.0 [16]. In this paper we report on our experiments to improve the coding performance and functionality such as scalability for low bitrate video applications while employing the DCT coding framework of the VM5.0.

The rest of this paper is organized as follows. In section 2, we present the basic coding structure employed in DCT coding framework of the MPEG-4 VM's. In section 3, we describe our proposed improvements to the DCT coding framework; these improvements have been recently accepted in the VM. In section 4, we present experimental results. In section 5, we summarize the main points of thi s paper.

2. CODING STRUCTURE

In MPEG-4 video coding [15,16] a scene to be coded is thought of as composed of several Video Objects (VOs) such that each Video Object can further be coded as one or more scalability layers referred to as Video Object Layers (VOLs). Furthermore, the time instances (snapshots) of a Video Object are referred to as Video Object Planes (VOPs). Generally speaking, VOPs are expected to be 2D mappings of VO's and thus considered to be of artbitary shape.VOPs are constructed by segmenting semantic objects of interest in each picture (frame) of a scene, thus one picture can be considered to be composed of one or more VOPs. If the entire scene is considered as one object and all VOPs are rectangular and of the same size as each picture then a VOP is identical to a picture. Similar to the fact that in MPEG-1 or MPEG-2, a picture may be coded as an I-picture, P-picture or B-picture, in MPEG-4, a VOP can be coded as a I-VOP, a P-VOP or a B-VOP. An I-VOP is intra-coded, a P-VOP is predictive-coded and a B-VOP is bidirectionally-predictive-coded. Thus although in MPEG-4 each VOP can be arbitrary shape and there is significantly more flexibility in coding, there is a basic correspondence in coding structure of MPEG-1, MPEG-2 and MPEG-4. Furthemore, for the case of rectangular VOPs of same size as each picture, the coding structure is practically identical.. For the case of rectangular VOPs, in Figure 1 we show an example prediction structure which is similar to M=2 structure used in MPEG-1 or MPEG-2 encoding. Such a structure is now possible due to recent introduction of B-VOPs.

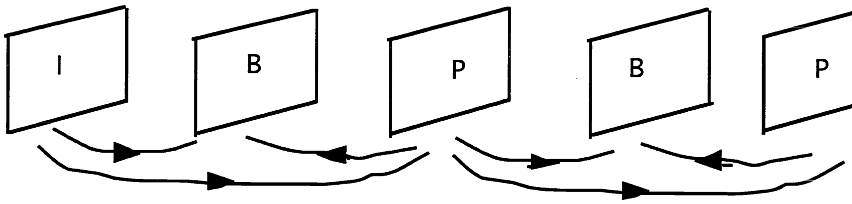

Figure 1 An example of I-, P- and B-VOP based coding now allowed in the MPEG-4 VM

The basic MPEG-4 VM coding currently employs motion compensation and DCT based coding. Each VOP is comprised of macroblocks that can be coded as intra- or as inter- macroblocks.The definition of a macroblock is exactly the same as in MPEG-1 and MPEG-2. In I-VOPs, only intra- macroblocks exist. In P-VOPs, intra as well as unidiretionall y predicted macroblocks can occur.where as in B-VOPs, both uni- or bidirectionally predicted- macroblocks can occur.

3. Coding of I-, P- and B-VOPs

We investigate three potential improvements. Our first experiment [11,12] is on improved coding efficiency of intra. Our second experiment [13] is the study of coding of B-VOPs to allow additional functionality such as temporal scalability at low bit-rates.

3.1 Intra-macroblock Coding in I- and P-VOPs

We introduce improved prediction of DC coefficients of intra DCT blocks. For reference, H.263 does not include DC prediction while MPEG-1 only allows a simple DC prediction. Then we introduce prediction of AC coefficients of intra Dct blocks, such prediction is not allowed in H.263 or MPEG-1. Further, we introduce block adaptive scanning of intra DCT coefficient blocks, again, this type of scanning is not included in H.263 or MPEG-1. Next, we propose use of a optimized VLC for intra DCT coefficient block coding, H.263 or MPEG-1 do not alow such VLC's, however, MPEG-2 allows them but MPEG-2 VLC structure is optimized for higher bitrates.

3.1.1 DC Prediction Improvement

The DC prediction method of MPEG-1 (or MPEG-2) is improved to allow adaptive selection of either the DC value of immediately previous block or that of the block immediately above it (in the previous row of blocks). This adaptive selection of the DC prediction direction does not incur any overhead as the decision is based on comparison of the horizontal and vertical DC value gradients around the block whose DC value is to be coded.

Figure 2 shows 4 surrounding blocks to the block whose DC value is to be coded. However, only three of the previous DC values are currently being used, the fourth value is anticipated to provide a better decision in the case of higher resolution images and may be used there. Assume, 'X', 'A', 'B', 'C and 'D' correspondingly refer to the current block, the previous block, the block above and to the left, the block immediately above, and the block above and to the right as shown.

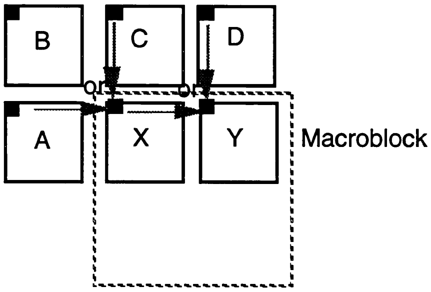

Figure 2 Previous neighboring blocks used in improved DC prediction

The DC value of 'X' is predicted by either the DC value of block 'A' or the DC value of the block 'C based on the comparison of horizontal and vertical gradients by use of Graham's method as follows.

The dc values 'dc' obtained after DCT are first quantized by 8 to generate 'DC' values.
$DC = dc//8$

```
if $(|DC_A - DC_B| < |DC_B - DC_C|)$
          $DC_{X'} = DC_C$
 else
          $DC_{X'} = DC_A$
```

For DC prediction the following simple rules are used:

- If any of the blocks A, B or C are outside of the VOP boundary, their DC values are assumed to take a value of 128 and are used to compute prediction values.
- In context of computing DC prediction for block 'X', if the absolute value of a horizontal gradient ($|DC_A - DC_B|$) is less than the absolute value of a vertical gradient ($|DC_B - DC_C|$), then the prediction is the DC value of block 'C', otherwise, DC value of block 'A' is used for prediction. This process is independently repeated for every block of a macroblock using appropriate immediately horizontally adjacent block 'A' and and immediately vertically adjacent block 'C'.
- DC predictions are performed identically for the luminance component as well as each of the two chrominance components.

3.1.2 AC Coefficient Prediction: Prediction of First Row or First Column

Either coefficients from the entire or part of the first row or the entire or part of the first column of a previous coded block are used to predict the co-located coefficients of the current block. For best results, the number of coefficients of a row or column and the precise location of these coefficients needs to be identified and adapted in coding different pictures and even within the same picture. This however results in either too much complexity or too much overhead. A practical solution is to use a predetermined number of coefficients for prediction, for example, we use 7 ac coefficients.

On a block basis, the best direction (from among horizontal and vertical directions) for DC coefficient prediction is also used to select the direction for AC coefficients prediction. An example of the process of AC coefficients prediction employed is shown in Figure 3.

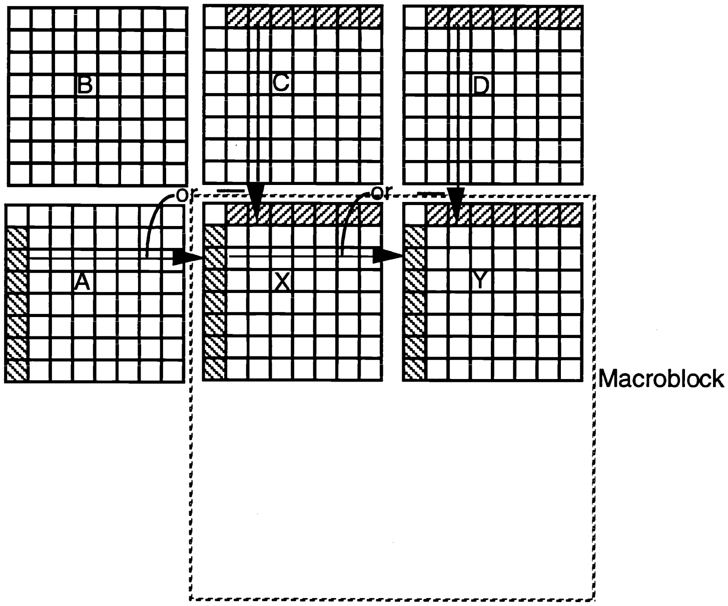

Figure 3 Previous neighboring blocks and coeficients used in improved AC prediction

Since, the improved AC prediction mainly employs prediction from either the horizontal or the vertical directions, whenever diagonal edges, coarse texture or combinations of horizontal and vertical edges occur, the AC prediction does not work very well and needs to be disabled. While ideally one would like to turn off AC prediction on a block basis, this generates too much overhead; thus we disable AC prediction at the macroblock basis. The criterion for AC prediction enable/disable is discussed in next.

In the cases when AC coefficient prediction results in a larger magnitude error signal as compared to the original signal, it is desirable to disable AC prediction. However, the overhead is excessive if AC prediction is switched on or off every block so AC prediction switching is performed ona macroblock basis.

If block 'A' was selected as the DC predictor for the block for which coefficient prediction is to be performed, we calculate a criterion, *S*, as follows.

$$S = \left(\sum_{i=1}^{7} | AC_{i0X} | - \sum_{i=1}^{7} | AC_{i0X} - AC_{i0A} | \right)$$

If block 'C' was selected as the DC predictor for the block for which coefficient prediction is to be performed, we calculate *S* as follows.

$$S = \left(\sum_{j=1}^{7} | AC_{0jX} | - \sum_{j=1}^{7} | AC_{0jX} - AC_{0jC} | \right)$$

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

Smarter legal research.