International Journal of Wavelets, Multiresolution
and Information Processing
Vol. 2, No. 2 (2004) 165–171

World Scientific
www.worldscientific.com

A FAST WAVELET-BASED VIDEO CODEC AND ITS APPLICATION IN AN IP VERSION 6-READY SERVERLESS VIDEOCONFERENCING SYSTEM

H. L. CYCON*, M. PALKOW†, T. C. SCHMIDT‡ and M. WÄHLISCH§

Fachhochschule für Technik und Wirtschaft Berlin
University of Applied Sciences
Treskowallee 8, 10318 Berlin, Germany
**hcycon@fhtw-berlin.de*
†*mpalkow@fhtw-berlin.de*
‡*schmidt@fhtw-berlin.de*
§*mw@fhtw-berlin.de*

D. MARPE

Fraunhofer-Institute for Telecommunications — HHI
Image Processing Department
Einsteinufer 37, 10587 Berlin, Germany
marpe@hhi.fraunhofer.de

The purpose of this paper is twofold: On the one hand, we propose a fast wavelet-based video codec which is implemented into a real-time video conferencing tool. The proposed codec uses temporal frame difference coding, a computationally low-complex 5/3 tap wavelet transform, and a fast entropy coding scheme based on Golomb–Rice codes. On the other hand, we present an application of the video conferencing tool in a serverless peer-to-peer IP-based communication framework. For mobile communication we propose a simple, ready-to-use location scheme for video conference users in a global network.

AMS Subject Classification: 94A08, 94A11, 94A45

1. Introduction

First, we propose a wavelet-based video codec specifically designed for simultaneous real-time encoding and decoding. Effective video codecs are usually designed as transform coding schemes extended by a prediction loop for exploiting temporal redundancies. The transform coder consists of three modules: an energy-concentrating, decorrelating transform module, a quantizer module controlling the rate-distortion (R-D) trade-off and an entropy coding module diminishing the redundancy of the quantized transform coefficients. Since a real-time video codec in a synchronous application has to encode and decode at least 25 frames per second (fps) simultaneously, there is an urgent need for low-complexity algorithms, at least

within software-based solutions. The main complexity bottlenecks for video codecs are usually the motion estimation in the temporal prediction loop, the transform module and the entropy coder module. The codec we describe in our presented approach uses frame difference coding only, i.e. we restrain from motion compensating techniques in favor of pure temporal frame differencing in order to limit the computational complexity on the encoder side. We also use a fast wavelet transform with short 5/3-tap filters additionally enhanced by a MMX-based[a] lifting scheme implementation. In order to speed up the third complexity bottleneck, we use an entropy coding scheme based on Golomb–Rice codes.[12] This scheme relies on a collection of codes related to different model probability distributions out of which the most suitable code will be selected to fit the probability distribution of symbols to be encoded. The R-D performance of the proposed codec is comparable with MPEG-4 or H.263 implementations operating at frame rates of 25 CIF frames per second when comparing typical head-and-shoulders video conferencing scenes and switching motion estimation off for both reference systems.

The codec described above is designed for real-time applications and has already been integrated into a video conferencing system. This system is implemented as an easy-to-use desktop video conferencing software running on ordinary desktop PCs or laptops using internet connections without a central Multi-Conference Unit (MCU). It is equipped with an innovative addressing system for locating mobile users and ready for next generation internet protocol version 6 (IPv6).

The paper is organized as follows. Section 2 contains a more detailed description of the real-time video codec and Sec. 3 deals with its application in a video conferencing tool. In Sec. 4 we introduce the idea of a distributed framework for mobile user location. Finally, in Sec. 5 we give some conclusions and an outlook for future experiments and developments.

2. Wavelet-Based Real-Time Video Codec

2.1. *Overview of the video codec*

Our coding scheme is basically a transform coder together with a simple frame-based temporal prediction loop as shown in Fig. 1. The transform coder consists of the reversible discrete wavelet transform (DWT) which decorrelates the signal, a quantizer (Quant.), and a lossless pre-coding/entropy coding step which compacts the data produced by the quantizer. For exploiting the temporal redundancy in a video sequence, only the residual signal between the current frame and the previous reconstructed one will be coded in the transform coding step, as shown in Fig. 1. Due to the linearity of wavelet transforms we compute the residual frame in the wavelet transform domain instead of calculating the frame difference in the spatial domain. This saves an inverse transform step in the temporal prediction loop.

[a]MMX denotes the "Multi Media eXtension" instruction set for Intel processors enabling a kind of Single Instruction, Multiple Data (SIMD) parallelism.

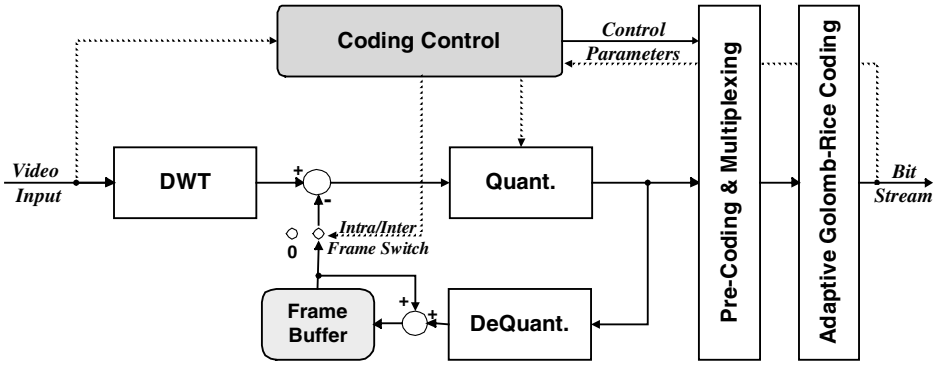

Fig. 1. Block diagram of our proposed video coding scheme.

Since we do not employ motion estimation techniques a periodic intra frame fresh-up after 99 inter, i.e. differentially coded frames is installed in order to avoid a too fast fading of image quality. However, this inter-to-intra frame relation can be manually chosen depending on the dynamical character of the video sources to be encoded.

In order to guarantee a constant transmitted bit rate on the average, we have implemented a backward operating coding control mechanism. We use a ring buffer with a dynamic level control which measures the quantized data stream. The bit rate can be adjusted on the fly by using control parameters during encoding. Note that these parameter changes have to be signaled to the decoder and therefore are integrated into the encoded bit stream.

Note also that this codec is a stripped-down speed-up version derived from conceptual ideas presented in Refs. 5 and 6.

2.2. *Transform and quantizer*

The transform module consists of a fast, computationally low-complex wavelet transform. We use wavelet transforms since they enable a transformation of each video frame as a whole without introducing blocking artifacts, especially at lower bit rates. By doing so, there is no need for a block decomposition of frames in the temporal prediction loop either. Our implementation of the wavelet transform uses 5/3-tap filters[14] in a computationally undemanding lifting scheme realization.[11]

As a quantizer module, we chose a simple uniform scalar quantizer with an enlarged dead zone. Note that the implementation of the quantizer is simultaneously integrated in the wavelet transform filtering by MMX techniques. This speeds up the coding process considerably.

2.3. *Pre-coding and Golomb–Rice entropy coding*

The data generated by the quantizer module are highly redundant. They have to be losslessly compressed by using an entropy coding module. The most effective

entropy coders in terms of minimum-redundancy are given by arithmetic coders.[12] For complexity reasons, however, we use an entropy coder consisting of a simple run-length pre-coder and a variable-length coder based on a set of Golomb–Rice codes. This scheme exploits similar techniques as have already been established in lossless still image coding.[13]

The quantized subbands are all treated individually since the statistical distribution in the different subbands are usually different. Especially at lower rates, there are also large numbers of concatenated chains of zero quantized transform coefficients, which the pre-coder separates from the bulk of the data by the means of a run-length encoding scheme. The significant (nonzero) coefficients as well as the run-length encoded insignificant coefficient data are encoded by using a family of Golomb–Rice codes.

2.4. *Performance evaluation of the video codec*

In native implementations, the video codec encodes and decodes 25 CIF frames (352×288 samples) simultaneously on a 500 MHz Pentium machine. Alternatively, five frames in PAL (720×576) resolution may be processed on the same platform. The image quality is comparable with MPEG-4 or H.263 coders when used in a frame difference coding mode only. At moderate motion complexity (for example head and shoulders) and at frame rates of about 25 fps in QCIF format (176×144) our coder produces a bit rate of approximately 200 kbit/s while delivering good visual quality.

3. The daViKo Videoconferencing System

The above proposed codec is the core of the digital audio–visual conferencing system daViKo[4] developed by the authors. The daViKo system represents a multipoint video conferencing solution without using a MCU. It has been designed in a peer-to-peer model (i.e. direct client-to-client communication) as an internet conferencing tool aimed at email-level use. Guided by the principle of serverless communication, daViKo refrained from implementing H.323 client requirements.[1]

By controlling the coding parameters appropriately, the software permits scaling in bandwidths from 64 to 4000 kbit/s on the fly. In addition, the encoded video resolution can be changed from sub-QCIF formats to PAL in ongoing sessions. Audio data are compressed using a MP3 algorithm with latencies below 120 ms depending on the chosen buffer size. Audio and video streams can be transmitted as unicast as well as multicast streams. An application-sharing facility is included for collaboration and teleteaching.

Due to low bandwidth requirements, the daViKo system is well suited to long-distance video conferences on a best effort basis. To strengthen its global usability, even on mobile devices, the user location scheme described below has been implemented into the system as well as advanced IPv6 network capabilities.

4. A Location Scheme for Mobile Users

Video conferencing is a synchronous form of communication requesting online presence of the participants. To retrieve the information on how to direct data flows to the appropriate user, that person's current device address needs to be resolved somehow. As device addresses change with mobility and as users may move between devices, a static address selection or any out-of-band information on user's presence are inappropriate.

Instead, a dynamic user session recording has proven advantageous. In the system introduced here, we denote this by a User Session Locator (USL) and store appropriate session information in an LDAP (Lightweight Directory Access Protocol) directory server. The video conference clients update information about ongoing sessions regularly, so that outdated session records can be identified by their timestamps. The USL server can be arranged within a local infrastructure not only to enhance scalability by distribution, but also to adopt local knowledge of the identity of users as well as a method for authentication. The global user look-up problem thereby reduces to deciding on unique user addressing and discovering the appropriate directory server for a given address. Current solutions either concentrate on a centralized directory as do MS Net-Meeting with the MS Internet Locator Server[9] or perform an Internet-wide user-based routing as is the purpose of the SIP server infrastructure.[3]

Our system *restricts* user addressing to email addresses because of its internet wide uniform availability, its convenience and ease of use. In adopting this restriction we radically break with telephone compatibility.

The Internet email system itself provides a mechanism for resolving user location through its interaction with the Domain Name System via the MX (Mail eXchange) record type for referencing a mail exchanger. Employing this commonly available Internet infrastructure we chose a simple strategy for locating a user's session directory. DNS data provided today are ready to cope with it: because the mail exchange record indicates a physically present domain where any requested user is identifiable along with a method of authentication, it is the appropriate location for a USL server. Within this domain, the look-up server can be identified by the common approach of a naming convention, i.e. usl.<mailexchanger-domain> .[8] Consequently, a global user look-up proceeds in two steps. Firstly, the MX record for the target user is requested, and secondly, the directory server hostname formed from the above naming convention is resolved (see Fig. 2).

Though simple, this user session information architecture neither relies on infrastructural changes nor requires dedicated user knowledge on the application side. From the mobility point of view, the USL servers play the role of distributed user home agents. Note that in contrast to H.323 gatekeepers or SIP servers the USL server consists of a passive session record store and can be realized by an unmodified standard LDAP server such as Open LDAP. For more detailed reading refer to Ref. 7. In a forthcoming paper,[10]

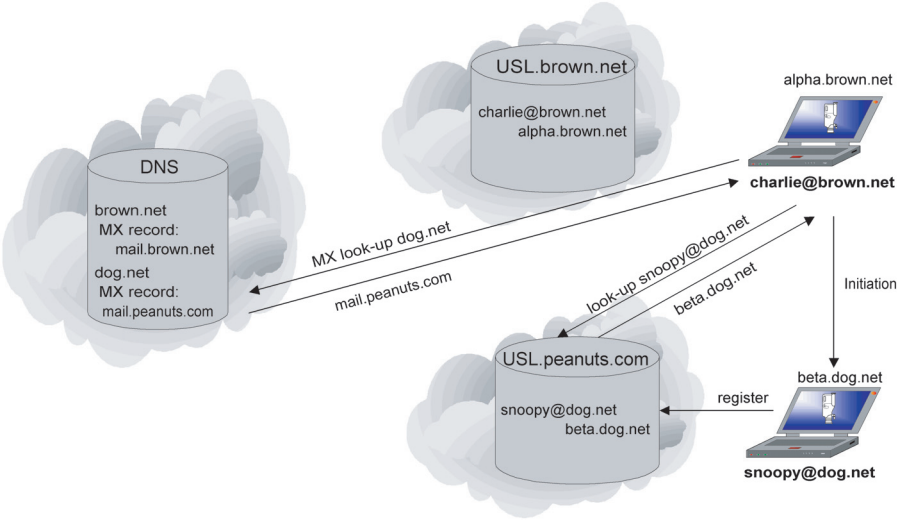

Fig. 2. Distributed user location scheme.

we will report on the problem of conference session preservation under device mobility.

5. Conclusions

A fast real-time wavelet-based video coding/decoding algorithm has been introduced. It is a simple frame difference coding approach using a low complex 5/3 tap wavelet transform, scalar quantization and a fast Golomb–Rice entropy coder. The proposed video codec is implemented in a video conferencing system for communication over IP-based networks using no central H.323-like MCU video server.

We also proposed a communication framework for its use within the video conferencing system. It is based on an innovative global addressing scheme and is currently being tested at the university FHTW (Fach Hochschule fuer Technik und Wirtschaft, University of Applied Sciences) Berlin. First experiences shows a growing acceptance by use and further ongoing developments aim at integrating the system into the wireless FHTW campus infrastructure.

References

1. ITU-T Recommendation H.323: “Infrastructure of audio-visual services — Systems and terminal equipment for audio-visual services: Packet-based multimedia communications systems”, Draft Version 4, 2000.
2. “The ISABEL Homepage”, http://isabel.dit.upm.es, 2002.
3. M. Handley, H. Schulzrinne, E. Schooler and J. Rosenberg, “SIP: Session Initiation Protocol”, RFC2543, March 1999.
4. “The daViKo homepage”, http://www.daviko.com, 2002.

5. D. Marpe and H. L. Cycon, Efficient pre-coding techniques for wavelet-based image compression, 1997, *Proc. PCS '97*, pp. 45–50.
6. D. Marpe and H. L. Cycon, Very low bit-rate video coding using wavelet-based techniques, *IEEE Trans. Circ. Sys. for Video Techn.* **9**(1) (1999) 85–94.
7. T. C. Schmidt, M. Wählisch, H. L. Cycon and M. Palkow, Global server less videoconferencing over IP, *Future Generation Comput. Syst.* **19** (2003) 219–227.
8. M. Hamilton and R. Wright, Use of DNS aliases for network services, RFC2219, October 1997.
9. NetMeeting Resource Kit Contents, Chap. 3, "Finding People" http://www.microsoft.com/Windows/NetMeeting/Corp/ResKit/Chapter3/default.asp, 2002.
10. T. C. Schmidt, M. Wählisch, H. L. Cycon and M. Palkow, Mobility aspects in IPv6 videoconferencing, in preparation.
11. W. Sweldens, The lifting scheme: A custom-design construction of biorthogonal wavelets, *Technical Report 1994:7*, Industrial Mathematics Initiative, Department of Mathematics, University of South Carolina, 1994.
12. A. Moffat and A. Turpin, *Compression and Coding Algorithms* (Kluwer, 2002).
13. M. Weinberger, G. Seroussi and G. Sapiro, The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS, *IEEE Trans. Image Processing* **9** (2000) 1309–1324.
14. A. Cohen, I. Daubechies and J.-C. Feauveau, Biorthogonal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* **45** (1992) 485–560.