User friendly H.264 for Realtime Editing

Research Planning Assignment 3

Pengpeng Ni
Department of Computer Science and Electronics
Mälardalen University

2005-11-07

Contents

1 Introduction

1.1 Definitions

- **Macroblock** A 16 x 16 pixel region within a video frame.
- **Inter prediction** Inter prediction is a basic coding technique which remove data redundancy between different video frames in temporal domain.
- **Intra prediction** Intra prediction is a basic coding technique which remove data redundancy in a single video frame in spatial domain.
- **I frame** I frame is a self-contained frame, encoded using only intra prediction.
- **P frame** P frame is encoded using inter prediction. Only previous decoded frame is used as referenced picture.
- **B frame** B frame is encoded using inter prediction. Not only previous decoded frame, but also the frame after the current frame in the display order are used as reference pictures.
- **Group Of Pictures** A Group of Pictures is formed by an I frame, together with all of the frames before the next I frame.
- **SP frame** A SP frame makes use of inter prediction from previously-decoded reference pictures, using at most one motion vector and reference picture to predict the sample values of each macroblock. Unlike P frame, the SP frame can be coded such that its decoded samples can be constructed identically to another SP frame or a SI frame.
- **SI frame** A SI frame is coded using intra prediction in a similar manner to I frame. It can be coded such that its decoded samples can be constructed identically to a SP frame.

1.2 Project specification

Our research project is aimed to support effective and quick browsing multimedia content over network. To enable user friendly browsing, a key technique is to provide full Video Cassette Recording (VCR) functionality. The set of full VCR functionality includes forward, reverse, step-forward, step-backward, fast-forward, fast-reverse, random access, pause, and stop. Among them, the reverse, fast-forward/reverse and random access are the trick operations that bring the problem.

Normally, the play rate of forward and reverse operations is defined as 25 fps (PAL/-SECAM) or 30 fps(NTSC). The fast-forward and fast-reverse operations put a speed up factor on the normal play rate. Intuitively, to get a perceived smooth playback effect, the latency between each two sequential displayed picture is supposed to be equally long. Given this assumption, there are two approaches to realize those fast play mode operations. One is to display each frame faster than the normal rate.Surpposing the normal play rate is 25 fps, the deadline assigned to the pipeline of transmission and decoding each frame is 40 ms. In other word, each frame gets 40 ms to get ready for display. With the speed up factor k = 10,the deadline assigned to each frame becomes 4 ms which is very short for most of client machines to perform the decoding task.In professional video editing application, a speed factor up to 30 is desired. The deadline can be 0.1 ms in the worst case. That will require the client machine to operate in an extremely high speed, which is not feasible. Therefor we consider frame dropping as the other approach and define the specification of the fast playback operation as follows, Given speed up factor k, to achieve smooth playback effect, only one out of k sequential frames will be displayed. Finally, we define random-access operation by jumping an arbitrary distance from the current displayed frame.

Specifically, we apply our research work to the video streaming system illustrated by Fig. 1. The multimedia content is MPEG encoded and stored as remote files in the center storage. The clients with the right to access to the MPEG files can view the video content while the video being streamed over the network.

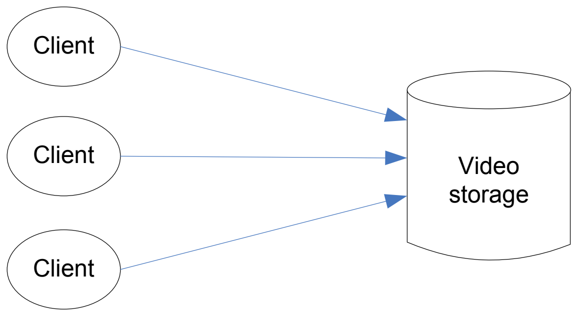

Figure 1: Video streaming system

1.3 Problem description

There are typically three major frame types defined in video compression standards. The three major frame types are referred as I(ntra) frames,P(redictive) frames and B(i-directional) frames. It is generally that a MPEG stream follows an I-B-P frame structure. An I frames followed by several P frames and B frames composes a Group of Picture (GOP). MPEG streams are comprised of a sequence of GOPs. Each P frame and B frame can be direct or indirect decoded dependent on the I frame at the beginning of a GOP so that it is usually to regard a GOP as an independent segment within a MPEG stream.

The I-B-P structure implies a typical forward dependence between sequential frames.This makes implementation of forward playback operation at normal play rate straightforward. However,the realization of the other operations like fast playback presents several challenges, such as huge memory requirement and higher network bandwidth. For instance, the straightforward implementation of reverse playback operation is to decode the whole GOP, store all the decoded frames in a large buffer, and display the decoded frames in the reverse order. However, this solution requires a huge buffer to store all of the decoded frames, i.e. if GOP size is N, a N-picture buffer is required and each picture may need at least several mega bytes. Such a large buffer is not feasible to a client station, especially when the GOP size is large. Moreover, the network traffic is increased heavily when a P frame or a B frame is requested in reverse direction or as random access point. The network needs sends all of the previous frames in the same GOP at a much higher rate which can be many times of that required by the normal forward-play operation. The number of those related frames can be up to the GOP size. The client needs also decode all of those frames at a much higher speed to minimize the latency of displaying a single frame. This requires much higher bandwidth of the network, higher performance and computation complexity in the client decoder, which is not realistic in most of cases.

1.4 Research method

We use deductive method to formulate the hypotheses to process the video streams. The background theory is in principle the existing MPEG coding standards and some mathematical transform method. Since the final product will be a measurement of human perceptive quality. The experimental result is supposed to be evaluated by qualitative methods.

Measuring visual quality can be both subjective and objective. Using objective criteria gives accurate, repeatable results so that objective quality measures are heavily relied by most of the video compression and video processing system. The most widely used measure is Peak Signal to Noise Ratio (PSNR), shown by Equation as follows. PSNR is measured on a logarithmic scale and depends on the mean squared error (MSE) of between an original and an uncompressed image or video frame

.

$$PSNRdB = 10 * log_{10}(2n - 1)^2/MSE$$

PSNR can be calculated quickly but it does not correlate well with subjective video quality. So some subject measures are required to complement the evaluation. We could make it by the form of questionnaire survey to different human observer, given a well built prototype demonstration.

2 Overview of the research area

In this section we give first a brief introduction of the compression techniques used in the video coding process,then we focus on the target video coding standard and give a detailed description of the interesting features which will be tightly related to our project.

2.1 Central literature

The central literature are the books referenced by [8],[2],[13]. The most interesting seminal papers are [6], [9].

2.2 Video compression background

Redundancy is defined in [2] as components that are not necessary for faithful reproduction of the piece of data. In real world, many types of data are very much *statistically* redundant, i.e. there is quite often little variation in the content of the background image in a video sequence, hence there is significant redundancy. Analysis of these statistical redundancy can allow the same information to be represented much more concisely. Hence, data compression is achieved by removing *redundancy*. Data compression techniques benefit mainly for the data storage and transmission systems. .

Video compression is a process combining different techniques in order to compact a digital video sequence into a smaller number of bits. These video compression techniques operate together by removing redundancy in the temporal, spatial, and/or frequency domains. A complementary process involves a compressor (encoder) and a decompressor (decoder).The encoder/decoder pair is often described as CODEC. A video CODEC encodes a source image or video sequence into a compressed form and decodes this to produce a copy or approximation of the source sequence. We illustrate a simplified block diagram of the video encoder by Fig. 2. A decoder can be simply regarded as a reverse process of the encoder. The goal of video compression is to achieve efficient compression while minimizing the distortion introduced by the compression process.

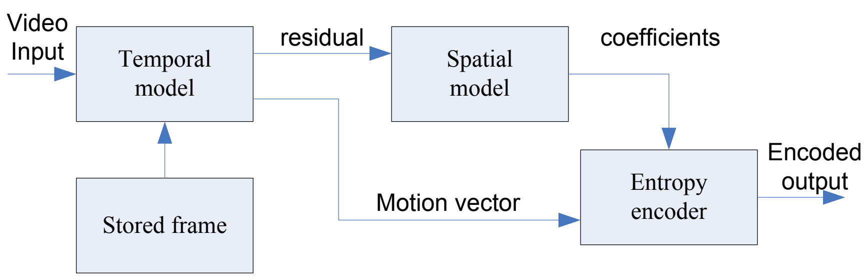

Figure 2: Simplied video encoder block diagram

In the temporal domain, there is usually a high correlation between images that were captured by video camera at around the same time,i.e. there is usually little change between two sequential video frames. The simplest method of temporal compression is to use the previous frame of two successive frames as the predictor for the current frame, and form a residual(difference) frame by subtracting its predicted frame from the current frame. Given the predicted frame, it is able to reconstruct the current frame using the residual frame. A residual frame usually contains small values with little variations comparing with what the current frame has. This make data compression possible because that data with smaller values can be represented by much fewer bits. The better the predicted frame matches the current frame, the smaller amount of bits is required to represent the residual frame. Hence, higher compression can be achieved by further exploration of the similarity between video frames.

A practical technique called *block-based motion compensation* is widely used to improve the accuracy of the prediction. The idea is derived from the fact that changes between video frames may be caused by object movements between the two frames. A better prediction may be formed by compensation for motions between the two frames. The block-based motion compensation technique divides first a video frame into several rectangular regions, for instance, a 16 x 16-pixel region within a frame, denoted as *macroblock*,is generally used as the basic unit of motion compensation in most of the current video coding standards. Motion compensation on macroblock involves two stages. At the first stage, motion estimation is conducted though searching an area in a previously-encoded picture(reference picture) centered on the current block position. A macroblock within the search area which gives the best match is chosen as the predicted macroblock and a vector describing the position of the selected macroblock relative to the current macroblock position is recorded. The vector is denoted as *motion vector* . At the compensation stage, the predicted macroblock is subtracted from the current macroblock to produce a residual macroblock which is encoded and transmitted together with the corresponding *motion vector*. The block-based motion compensation provides reasonable effective temporal compression so that it has become the basis of the temporal compression used by all current video coding standards. Motion compensation is also known as inter prediction. Video frames decoded using inter prediction from previously decoded frames is referred as P(redictive) frame. A P frame can be predicted by one or several previously decoded frames which are denoted as *reference pictures*.

In the spatial domain, there is usually a high correlation between pixels that are close to each other, i.e. the values of the neighboring samples are often very similar. It is able to form a prediction block based on previously encoded and reconstructed blocks within the same current video frame. The spatial compression is also know as intra prediction. Video frames encoded using only intra prediction is referred as I(ntra) frame. It is self-extracting so that I frame can be used as access point of a video sequence.

The video compression in the frequency domain involves exploiting relationships between frequency of the variation in the data values. To convert image or motion-compensated residual data from spatial domain into frequency domain, we need use a transformation method, for example Discrete Cosine Transform(DCT), Discrete Wavelet Transform(Wavelet) and etc. The DCT can process the image in units of block, and decorrelate the image data in a data-independent manner. Because of this, it has gained tremendous popularity and been therefore the heart technique in the MPEG and JPEG standards. The DCT produces a 2D matrix of the frequency content corresponding to the spatial image data. The frequency matrix shows if the value belonged to the original matrix are changing rapidly or slowly. It helps also to separate the image into spectral sub-bands of different importance. Low frequency components reflects the features on the large scale which are more important with respect to the image's visual quality, while high frequency components implies the changes on a short distance scale. For most of images, much of their signal energy lies at low frequencies. Hence large values usually appear in the upper left corner of the DCT matrix and the lower right values (higher frequencies) are often small. Data compression is possible because high frequency components are small enough to be neglected with little visible distortion.

Generally, transform and motion compensation do not compress any data itself. The compression comes from *quantization* of the transform coefficients and the further processing work. Quantization is the heart of any lossy scheme. It is the quantization stage that introduces the distortion. A quantization algorithm map a set of input data to a single output data. The number of distinct values is then reduced to a much smaller set. For example, it is able to map large value to smaller value using a common divider. Then it is acceptable to only store lower frequency components in the transform coefficients and map the small high frequency components by zero. The essential information of an image can still be represented by the quantized value with small loss in the visual quality.

After the previously introduced processes, a video data stream is now represented by a serial symbols, such as the quantized transform coefficients and motion vectors. These symbols are prepared for the final stage of video compression, named as*Entropy encoding*. Entropy is a measure of the disorder of the system in science. The definition of entropy is aimed at identifying often-occurring symbols in the data stream as good candidates for short codewords in the compressed bitstream [2]. *Variable-length coding*(VLC) is the best-known *entropy coding* method. Basically, VLC assigns a codeword to each symbol according the frequency count of their occurrences. More frequent a symbol occurs, shorter codeword is assigned. The average length of the codewords is shorten by doing this. There is also another group of entropy coding method denoted as *arithmetic coding*. However, since the entropy coding is out of the range of our research, we don't give further introduction concerning on *entropy coding*.

2.2.1 H.264/AVC - the latest video coding standard

The video coding standard, H.264 is a joint project developed by the Video Coding Experts Group(VCEG) and the Moving Picture Experts Group(MPEG). It was standardized in 2003, published as document ISO/IEC MPEG-4-AVC. Concentrating specifically on efficient compression of video frames, H.264 uses network-friendly syntax and introduces several new encoding features which can be taken advantage of by different applications. Comparing with the previous coding standards, up about 50% bit rate saving can be achieved for a given video quality. The significant improvement in compression efficiency has gained H.264/AVC tremendous popularity in industries.

The video coding process of H.264/AVC is similar in spirit to other standards but with some important differences as follows:

Enhanced motion estimation with variable block size The minimum unit of motion compensation is not the macroblock consisting 16 x 16 samples. A macroblock can be further divided into smaller partitions such as 8 x 4 or 4 x 4 samples. This feature adapts to the variation in the size of moving objects.

Multiple reference pictures Not like the previous MPEG standards, P and B frames is predicted by more than one or two reference pictures. This increase the coding complexity but enables the encoder to search for the best match for a macroblock.

SP/SI frames H.264 has defined two new frame types, referred as SP and SI frames in its extended profile. Similar to P frame, SP frame makes use of inter prediction from previously-decoded reference pictures, using at most one motion vector and reference picture to predict the sample values of each block. Unlike P frame, the SP frame can be coded such that its decoded samples can be constructed identically to another SP frame or a SI frame. In other word, it allows identical reconstruction of a frame even when different reference pictures are being used. A SI frame is coded using intra prediction in a similar manner to I frame. It can be coded such that its decoded samples can be constructed identically to a SP frame. The properties of SP and SI frames aid in providing functionalities for bitstream switching, splicing, random access, fast-forward, fast-reverse and error resilience/recovery. A detailed introduction of the SP/SI design is given in [6],[7].

2.3 Related works

Different video coding methods have been presented to address the implementation of full or part of VCR functions for MPEG compatible video stream. We introduce some of these methods as follows.

2.3.1 Video reverse transcoding

Because MPEG video stream is encoded using forward motion compensation of MPEG video stream, it is the reverse-playback function that presents the major difficulty. A couple of previous research works focusing on reverse decoding problem have been done in the video transcoding area.

Chen et al. [1] presents a stream conversion transcoder that encodes P frames as I frames locally after the decompression and playout of each P frame. The original MPEG I-B-P is converted to a local I-B streams. In the I-B stream, the B frames can simply swap its forward and backward motion vectors to compensate the reordering of its predicted frames. [5]. The modified I-B stream can then emulates reverse play in a generic decoder. However, the disadvantage of this method is the complexity of the P to I conversion.

The concept of transforming incoming stream into a local stream can be realized in simpler ways. Wee et al. [4] presented a method to transform the incoming stream in the compressed domain. In other words, the transcoder process the incoming stream before the decompression. Several reverse-play transcoding algorithms have been presented and compared in [4], [5]. Those algorithms estimate the reverse motion vector based only on the forward motion vector. Thus the time-consuming procedures of motion compensation and/or intra prediction are skipped, which reduce considerable computation complexity of the transcoding process. Those transcoding algorithms trade off degrees of computational efficiency for motion vector accuracy, therefore the distortion caused by motion vector estimation is a big drawback.

The reverse transcoding method does not address the problem of extra network traffic caused by fast-forward/backward and random access. Moreover, it addresses the problem of reverse-play for video stream compliant to previous MPEG coding standards. In the previous MPEG coding standards, each P frame using only one reference frame. The latest MPEG standard H.264/AVC allows using more than one previously coded frame as a reference picture for P- and B-frames [3], [2], [13]. The algorithms need to be redesigned/demonstrated to suit the new features provided by H.264/AVC.

2.3.2 Video Streaming

Dual streams with least-cost selection

Lin et al. [9] presents a "dual bitstreams" solution to solve the full VCR functionality. It is proposed to offline re-encode an original MPEG video sequence in the reverse order to generate a secondary bitstream. The encoding can be arranged so that the I-frames in the backward bitstream are interleaved between I-frames in the forward bitstream. A server process is required to handle the client request. It can selects the nearest I-frame to the current displayed frame out from the two video streams and send the related frames between the nearest I-frame to the next requested frame to the client. Due to the interleaving of I frames in the dual streams, the GOP size is actually shortened. Therefor the actual number of frames to be sent and decoded can be significantly reduced.

SP and SI frames in video coding standard H.264

SP/SI frames were first proposed in 2001 by Karczewicz and Kurceren [6]. Since the SP frames allow identical frames to be reconstructed even when different reference frames are used. It can be used instead of I-frames as access points when switching between bitstreams. Several application scenarios of using SP/SI frames have been provided in ieee:SPDesign. We want also apply the SP/SI frames in our project. A example of how to use the SP/SI frames is given in section3.1.

2.4 Key conferences and leading research groups

2.4.1 Key conferences

The IEEE International Conference on Image Processing (ICIP)

ICIP is widely regarded as the premier forum for the presentation of technological advances and research results in the fields of theoretical, experimental, and applied image and video processing. It sponsored by the IEEE Signal Processing Society and has been held annually since 1994.

The areas of interest of ICIP 2006 include but are not limited to:

- Image/Video Coding and Transmission
- Image/Video Processing and Analysis
- Image Formation
- Image Scanning, Printing, Display and Color
- Image/Video Storage, Retrieval, and Multimedia
- Applications

ACM Multimedia

ACM Multimedia is the premier international technical multimedia conference attended by researchers from both academia and industry. The conference is held annually and covers all aspects of multimedia computing

The topics of interest includes context-aware multimedia communications, Internet telephony, peer-to-peer streaming, audio/video streaming, multimedia content distribution, wireless multimedia, adaptive support for scalable media, Internet protocols, multimedia servers, operating systems, middleware and QoS.

IEEE International Conference on Multimedia & Expo (ICME)

Similar as ACM Multimeida conference, ICME is a major annual international conference bringing together researchers, developers and practitioners from academia and industry working in all areas of multimedia. ICME provides the chance for exchange the results of state-of-the-art research, development, and implementations of multimedia systems, technologies, and applications. ICME also host industrial exhibitions during conference period.

The topics of interest are:

- Multimedia content management,includes Multimedia content analysis, indexing, and retrieval, Multimedia content modeling Multimedia content description, Multimedia standards and related issues, Media mining, Personal multimedia
- Multimedia signal processing, includes Signal processing for media analysis and integration, Media synchronization,Media filtering and enhancement and etc.
- Multimedia communications, includes Multimedia streaming, networking and distribution, Multimedia authentication, content protection and digital rights management, Multimedia QoS, Context-aware multimedia and etc.

2.4.2 Research groups

There are a lot of research group which are active in the multimedia area.

The Beakley Multimedia Research Center

They developed the first widely-distributed software decoder for MPEG-1 video in 1992, and a software encoder for MPEG-1 video 1993. Both the decoder and encoder are widely used in different applications, the code has even been used by dozens of companies to develop their products. The research group directed by Lawrence A.Rowe still work on their MPEG tools, and the current focus has been the development of perceptual coding models that can produce video stream with the highest perceived quality for a fixed bitrate.

Image,Video, and Multimidia Systems (IVMS) at Stanford university

The group is part of the information systems laboratory in the Department of Electrical Engineering in Standford university. Bernd Girod leads the research group. They carry out fundamental and applied research on various aspects of video compression, coding, and networked real-time media systems. Topics of interests are such as Rate-distortion optimization, adaptive video coding and etc. They have also done some analysis about H.264/AVC, see [12], [11]. Additionally, an interesting H.264 coder is developed by them [10].

The signal processing group in Chalmers

A Swedish research group, was established in 1988. Tomas Gustavsson is the leader of the group. They're working with interesting topics like object detection, recognition and classification. Signal processing for communications is also belongs to their current research subjects.

Research labs from industry

Large companies like IBM, Microsoft and HP have established their own labs to do research work on hot topics in multimedia area such as video content analysis and recognition, media streaming, image processing and etc.

3 Research plan

3.1 Hypothesis

It is possible to utilize the features of SP/SI-frames to support quick and user friendly browsing of multimedia content via networks. By carefully design the encoding parameters, a H.264 compliant SP/SI coding scheme is available which effectively trade off the responsibility of each requested frame with tolerable quality distortion.

3.2 Concrete research questions

The concrete research questions are related to time complexity, compression efficiency and rate distortion of designing the coding scheme. Meanwhile the SP/SI encoding scheme has the similar patten as what the dual streams scheme has, see [2.3.2]. Two encoding schemes are to be compared using those criterions. Additionally, the SP/SI design assumes the client has a feedback channel so that the server can send the proper frame according to the received feedbacks. The feasibility of the SP/SI solution to those client without feedback channel need also to be analyzed.

We have the concrete research questions as follows?

- What is the time complexity of the SP/SI coding process?
- How efficient is the SP/SI scheme to solve the VCR functions? What is the average number of frames need to be transmitted and decoded?
- How to measure the rate distortion introduced by SP and SI frames?
- How does the SP/SI frames compare to "dual streams" method?
- If it is possible to apply the SP/SI frames to some video streaming system not built on client-server mode?

3.3 Concrete expected results

The concrete expected results are as follows:

The time complexity, measured in terms of FLOPS for both SP/SI encode process and decode process,on the impact of different VCR functions, such as "play x 1", "play x -1", "play x 10","play x -10" and random access.

The compression efficiency, measured in bitrates. Distinguished from the "propagation speed", bitrate measures quality transmitted per time units. Higher bitrate implies higher compression efficiency. We want to know how much bitrates the SP/SI encoder can provide.

The expected results include also a measurement in terms of a usability versus rate distortion. Usability can be the average response time of the requested frame.

3.4 Research Activities

We have performed literature survey in the relative area. The stat of art solutions to our research problem are reviewed and compared. A video coding scheme is proposed to solve the problem. Focusing on its particular application, we have defined the aspects to be evaluated and how to perform those evaluation. We also analysis the feasibility of the SP/SI frames. A flexible interface between the client and server is in the ongoing discuss.

3.4.1 Evaluation

The goal of the encoding scheme is to improve the responsiveness of each required frame by reducing the coding complexity and bandwidth requirement. However, the SP/SI frame introduce some quality loss due to the additional quantization. There is a tradeoff between the responsiveness and the video quality. We evaluate first the tradeoff and therefore conclude the availability of the encoding scheme.

The rate distortion theory can be used to evaluate the performance of the coding scheme. It determines the minimal bitrate of a video stream with given distortion. The bitrate is usually referred as the number of bits that one second of video data will consume.The distortion is often defined as the variance of the difference between input and output signal. We formulate a theoretical analysis and then perform the experiment to verify the theoretical analysis.

Additionally, the number of the extra frames being decoded is proportional to the coding complexity. It is interesting to analysis the minimum, maximum and average number of total frames to be sent over the network under the impacts of the random access and fast playback operations. Those numbers can be used as one of the criterion of the responsiveness analysis.

Combining these analysis, we look forward to finding out suitable value of parameters such as quantization parameters and the proportions of SP and SI frames, in order to encode H.264 compliant video streams for the use of our application.

3.5 Planned publications

MPEG streams for user friendly editing

A state of art report, which provides background introduction of MPEG coding standard such as MPEG4 and H.264, a survey of state of art techniques supporting VCR functionality or reverse coding problem. In this paper, we analyze also the feasibility of the current and previous research work that have been done in the same field.

SP and SI frames supporting VCR functionality

In this paper, focusing on SP and SI frame of H.264/AVC, we describe how to encode SP and SI frames and demonstrate how to utilize these features to support full VCR functionality. We analyze the efficiency and complexity of a coding scheme using SP/SI frames and compare this method to the "dual-streams" coding method. This paper try to answer most of the concrete research questions in 3.2.

A tradeoff between display latency and rate distortion

We will evaluate the SP/SI coding scheme in terms of responsiveness and bitrate distortion. Response time analysis and bitrate distortion measurement will be involved in this paper.

Resource requirements analysis in a SP/SI codec

There are questions about resource requirement and management in video codec. According to the scheme described in 3.1, the SP/SI encoder stores a secondary representation of a video frame, and send it as a substitute to the primary representation in necessary situation, such as backward playback and error recovery. How to store the secondary SP/SI frame in order to improve the performance will be the topic in this paper.

User friendly codec for realtime editing

Assume we can hopefully get the positive result from the previous analysis work about the SP/SI coding scheme, a paper will be written covering the design and implementation results of a real H.264 codec for industry use.

3.6 Conference trips

One of our sponsors Company Ardendo AB usually attends the exhibitions organized by two international conferences each year. One is of National Association of Broadcasters (NAB), at April in Las Vegas. The other is of International Broadcasting Convention (IBC) at September in Amsterdam. There are many workshops and forums which can be interesting to attend.

We will hopefully visit some conferences in academic world, it depends on if we get article accepted. One of the interesting conferences is ICIP, which is generally regarded as the most premier conference in multimedia area.

3.7 Time plan and milestones

Spring 2005:

- Study the multimedia compression techniques and details of H.264, MPEG4 standards.
- literature survey, research area overview.
- workload from industry.

Autumn 2005:

- paper 1 (MPEG streams for user friendly editing).
- About 15 course credits will be obtained

Spring 2006:

- paper 2 (SP and SI frames supporting VCR functionality).

Autumn 2006:

- paper 3 (A trade off between display latency and bitrate distortion)
- Milestone: Thesis proposal for Licentiate degree.
- Total 30 points course credits will be obtained.

Spring 2007:

- Milestone: Licentiate Thesis defence

Autumn 2007:

- Paper4Resource requirements analysis in a SP/SI codec

Spring 2008:

- Paper5 (User friendly codec for realtime editing).

Autumn 2008:

- Ph.D thesis proposal;
- Total 50 course credits will be obtained

Spring 2009:

- Milestone: Ph.D thesis defence

References

[1] Ming-Syan Chen and Dilip D.Kandlur. Downloading and stream conversion:supporting interactive playout of videos in a client station. In *Proc.2nd Int.IEEE Conf.Multimedia Computing and Systems*, 1995.

[2] Iain E.G.Richardson. *H.264 and MPEG-4 Video compression - Video Coding for Next-generation Multimedia*. Wiley, first edition, 2004.

[3] ITU-T and ISO/IEC JTC 1. *Advanced Video Coding for Generic Audiovisual services, ITU-T Recommendation H.264*, April 2003. ISO/IEC 14496-10(AVC).

[4] Susie J.Wee. Reversing motion vector fields. In *Image Processing, 1998.ICIP 98.Proceeding*, volume 2.

[5] Susie J.Wee and Bhaskaran Vasudev. Compressed-domain reverse play of mpeg video streams. In *SPIE International Symposium on Voice,Video,and Data Communicatins*, November 1998.

[6] Marta Karczewicz and Ragip Kurceren. The sp- and si-frames design for h.264/avc. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7), July 2003.

[7] Ragip Kurceren and Marta Karczewicz. New macroblock modes for sp-frames. In *ITU-T Video Coding Experts Group Meeting*, December 2001.

[8] Ze-Nian Li and Mark S.Drew. *Fundamentals of Multimedia*. Pearson Prentice Hall, 2004.

[9] Chia-Wen Lin, Jian Zhou, Jeongnam Youn, and Ming-Ting Sun. Mpeg video streaming with vcr functionality. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(3), March 2001.

[10] Eric Setton. an sp frame encoder/decoder for the new h.264 standard. WWW Page, December 2004.

[11] Eric Setton and Bernd Girod. Rate-distortion analysis and streaming of sp and si frames. September 2001.

[12] Eric Setton and Bernd Girod. Video streaming with sp and si frames. In *Preceedings VCIP,Beijing*, July 2005.

[13] John Watkinson. *The MPEG handbook*. Focal Press, 2004.