Analysis of H.264 video encoded traffic

H. Koumaras, C. Skianis, G. Gardikis, A. Kourtis

Institute of Informatics and Telecommunications NCSR «DEMOKRITOS»,
Patriarchou Gregoriou Str., Agia Paraskevi, Attiki, 15310 Athens Greece
e-mail: {koumaras, skianis, gardikis, kourtis}@iit.demokritos.gr

Abstract

In future wireless and wired communication networks, video is expected to represent a large portion of the total traffic, given that multimedia services, and especially variable bit rate (VBR) coded video streams, are becoming increasingly popular. Consequently, traffic modeling and characterization of such video services is very important for the efficient traffic control and resource management of complex networks. The new H.264/AVC standard, proposed by the ITU-T Video Coding Expert Group (VCEG) and ISO/IEC Moving Pictures Expert Group (MPEG), is expected to dominate in upcoming multimedia services, due to the fact that it can achieve considerably higher coding efficiency compared to the already widely used standards. This paper presents both a frame and a layer (i.e. I, P and B frames) level analysis of H.264 encoded sources. Analysis of the data suggests that the video traffic can be considered as a stationary stochastic process with an autocorrelation function of exponentially fast decay and a marginal frame size distribution of approximately Gamma form.

Keywords

H.264 video coding, Traffic analysis, Video streaming

1   Introduction

Multimedia applications and services have already possessed a major portion of the today traffic over computer and mobile communication networks. Among the various types of multimedia, video services (transmission of moving images and sound) are proven dominant for present and future broadband networks.

The raw video data have very high bandwidth and storage requirements making its transmission and storage impractical and economically unaffordable. For this reason, a lot of research has been performed on developing compression techniques, which exploit both temporal and spatial redundancy in video sequences.

From the advent of video coding, two main encoding schemes were proposed and are still used: The Constant Bit Rate (CBR) and the Variable Bit Rate (VBR) modes. In VBR mode the quantization parameters are maintained constant for the encoding process. So, the deduced video quality is almost sustained steady but the derived encoding bit rate fluctuates around a mean value. On the contrary, in CBR mode a rate-control algorithm alters the quantization parameters dynamically according to the frame complexity, in order to achieve the required target bit rate.

By this way the picture quality is not maintained constant, but it changes reversely analogous to the spatial and temporal activity of each video frame.

A major surge of interest in the topic of VBR video traffic modeling has appeared with the advent of video coding. Relevant early studies examined various characteristics of VBR video traffic, such as differences in successive frame sizes and cluster lengths (Chin *et al*, 1989) or scene duration distributions (Verbiest *et al*, 1988). Also recently (Skianis *et al*, 2003), (Doulamis *et al*, 2000) introduced efficient modeling tools and techniques of VBR MPEG-1/H.261 coded video at frame and GOP level.

Results from these and other works indicate that the histogram of frame sizes exhibits a bell-shape (Heyman *et al*, 1992), (Maglaris *et al*, 1988), (Nomura *et al*, 1989). Furthermore, correlations in the video bit rate are found to decay exponentially (Cohen and Heyman, 1993), (Haskell, 1972), (Heyman *et al*, 1992), (Lucantoni *et al*, 1994), (Maglaris *et al*, 1988), while other studies (Nomura *et al*, 1989), (Ramamurthyi and Sengupta, 1990), (Rodriguez-Dagnino *et al*, 1991) observe a more complex phenomenon, in which the correlation decay is rapid for the initial lags, then continues at a lower rate.

The most popular and widely used encoding algorithms are the ones developed by the Moving Picture Experts Group (MPEG) and the Video Coding Expert Group (VCEG) of the ITU. Recently these two organizations jointly developed a new codec, the H.264 or MPEG-4 Part 10 Advanced Video Coding (AVC) codec (Wiegand *et al*, 2003). Featuring updated capabilities, the new codec can achieve 40-50% compression efficiency gain over today's optimized MPEG-2 codecs. Due to the advances of H.264, it is expected that it will prevail in future networks and mobile application systems, making traffic modeling and characterization of H.264 video streams a useful tool for network managers and designers. In this respect, this paper studies the output traffic from an H.264 codec at the frame and the layer level. In particular, the work focuses on video traffic that is generated by VBR H.264 coding since it offers relatively constant quality and has less bandwidth and storage capacity requirements.

This paper presents a frame and layer (i.e. I, P and B frames) level analysis of H.264 video streams, showing that H.264 video traffic can be described as a stationary stochastic process with an autocorrelation function of exponentially fast decay and a gamma-like marginal frame size distribution.

The rest of the paper is organized as follows: Section 2 presents the new characteristics and enhancements of the H.264 standard, Section 3 discusses on the statistical analysis of the H.264 video stream and finally Section 4 concludes the paper.

2 The Essential Enhancements of the new H.264/AVC standard

In 1998 the ITU-T VCEG issued a call for proposals (H.26L project), with main scope to double the coding efficiency in comparison to the already existing coding standards. In 2001, VCEG and ISO/IEC MPEG formed a Joint Video Team (JVT) in order to finalize the standard and to submit for formal approval as H.264/AVC (Wiegand *et al*, 2003).

The new standard included many enhancements in the coding process, which contribute to the improved coding efficiency of H.264/AVC. Some essential indicative enhancements are:

- Variable block size support for motion compensation with luma block sizes down to 4x4, in conjunction with 4x4 level transformations. (see Figure 1).
- Quarter-sample motion vector accuracy.
- Extended reference frame selection for P frames, among various previously decoded frames.
- De-blocking filter within the motion-compensated prediction loop.
- New context-based adapted entropy coding methods: CAVLC and CABAC.
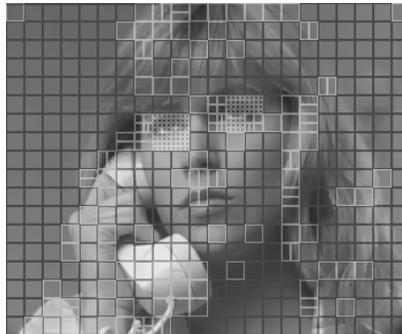
Figure 1. Example of the variable block size approach of H.264/AVC

The main target of the aforementioned enhancements is the perceived quality improvement and the high-compression efficiency. In this respect, due to the expected business models in emerging wired and wireless networks, where the end-user costs are relative to the transmitted data volume, the bandwidth occupation and utilization, compression efficiency is core goal for all the future multimedia services. For these reasons, H.264/AVC is expected to dominate in future wireless and wired networks.

3 Statistical Analysis of the H.264 encoded data

3.1 Frame Level Analysis

For the statistical analysis of H.264/AVC encoded data, the reference encoder JM is used, considering encodings without rate control and fixed quantization parameters for all test sequences. In H.264, the three common different frame modes are adopted, namely: Intra-frame (I), Predictive (P) and Bidirectional predictive (B), widely referred as I, P and B. In particular, the I frames are also called Intra frames, while B and P are known as Inter frames. The combination of successive types of frames forms a Group Of Pictures (GOP), whose length is mainly described by the distance of two successive I frames. In the described work, frame rate is set constant at 25fps, coding GOP structure is set as IPBPBPBPB… and Intra-period adopts values between 3 and 12 frames.

In this respect, Figure 2 illustrates the size of 1100 frames of an H.264 test signal (encoded with quantization scale 20 for all the frames and GOP length 12), where it can be noticed that the large frame sizes (periodical peaks in the figure) correspond to I frames, while the smaller ones are B frames and the intermediate frame sizes are P frames. Moreover, the periodicity that seems to

appear in the peaks of I frames, corresponds to the distance of two successive I frames, which reveals the length of the used GOP. It is also noted that the frame size follows the spatial and temporal activity of the test signal, where more complex frames require more bits for their description, while static and simple frames are described by fewer bits.

Also another interesting observation is that inter-frames (i.e. P and B) present more intense fluctuation in comparison with the Intra frames. This stems from the fact that according to the content dynamics of the video signal, some Macro-Blocks (MBs) of the inter-frames may be intra-coded, which results in lower compression ratio and therefore higher frame sizes. Figure 3 depicts the total number of Intra MBs for the inter-frames of the 1100 frames of Figure 2. It can be observed that the shape of the Intra MBs vs. Inter-frames graph (Figure 3) plays a major role in the form of the frame size graph (Figure 2). In other words, inter-frames appear to influence largely the actual video traffic.
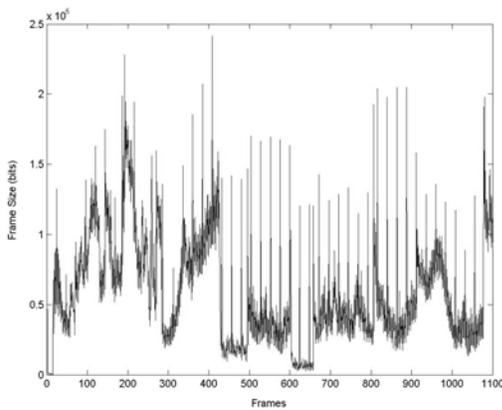
Figure 2. The frame level analysis over a time-window of 1100 frames
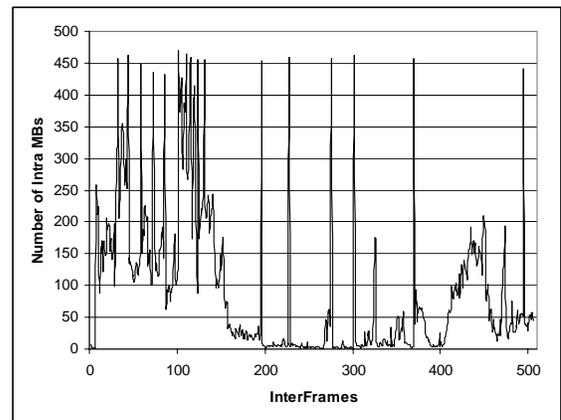
Figure 3. The total number of Intra MBs for the inter-frames over a time- window of 1100 frames
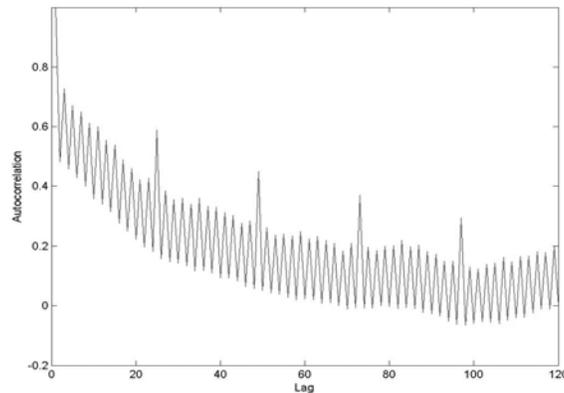
Figure 4. The autocorrelation of the 1100 frame sizes

Figure 4 illustrates the autocorrelation function for the 1100 frames. It can be observed that the autocorrelation graph consist of periodic spikes that are superimposed on a decaying curve. The

highest peaks correspond to the autocorrelation of the Intra frames of the video sequence, which are followed by 11 lower spikes before the next "Intra" peak. This periodicity reveals GOP length of the sequence. The lower spikes between two successive "Intra" peaks correspond to P frames, which are typically smaller than the I-frames. Finally, the wells between I and P peaks, correspond to the B frames of the test sequence, which are the smaller frames of all.

Based on the already discussed results, it can be deduced that the behavior of the H.264 encoded signal can be described as a superimposition of the three different distributions, which result from three different frames modes (i.e. I/B/P). Therefore, elaborating each frame type separately is more efficient and produces more detailed description of the H.264 video traffic. The next section presents an I/B/P layer analysis of the encoded signal.

3.2   I/B/P Level Analysis

In this section, the analysis is performed at I/P/B level. In this respect a video segment from the film "Spider-man II" is used as the reference signal. This segment consists of 18357 frames of YUV 4:2:0 format in 528x384 resolution, where encoding is performed using JM H.264 reference encoder, at VBR mode with constant GOP structure of the form IPBPBPBPB…. In order to study the nature of the video stream, intra-frame period and quantization parameters are altered during the experiments. During each encoding process, video traces are captured, containing data on the type and the size of each encoded frame. As a result, frame statistics based on specific quantization scale and encoding settings are derived and depicted in Table 1 in the form of mean values and variances of I/P/B frame sizes. The notation (x,y,z)-l is used for the quantization scales of I,B,P frames and the selected intra-frame period (i.e. GOP length).

| Quantization Settings / Frame Types	I Frames		B Frames		P Frames	
	Mean (10^5 bits)	Variance (10^9 bits)	Mean (10^5 bits)	Variation (10^9 bits)	Mean (10^5 bits)	Variation (10^9 bits)
(10,10,10)-12	3.5447	7.6209	2.2765	3.3255	2.7134	4.3131
(20,20,20)-12	1.4816	3.3937	0.4302	1.1001	0.6701	1.7616
(30,30,30)-12	0.5391	0.6596	0.0786	0.0759	0.1633	0.1940
(20,20,20)-3	1.4721	3.2852	0.4286	1.1060	0.6722	1.7777
(20,20,20)-6	1.4846	3.3674	0.4310	1.1086	0.6706	1.7707

Table 1: Frame statistics overview of the encoded signal

From Table 1, it can be derived that higher encoding parameters, which cause coarser encoding quality, result in lower mean frame sizes and variations in comparison with lower quantization parameters, which produce better encoding quality. On the contrary, the alternation of Intra-frame period does not affect frame sizes, which remain practically constant.
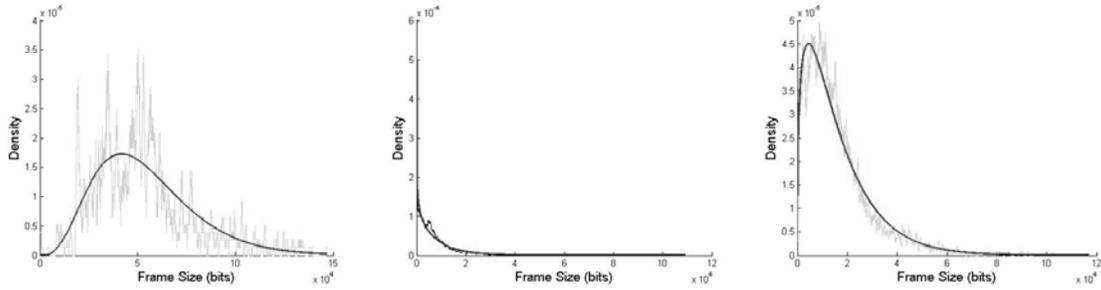In order to study the statistical behavior of the encoding stream, the Probability Density Functions (PDFs) for each frame type of the encoded signal at various quantization scales are drawn in Figure 5.
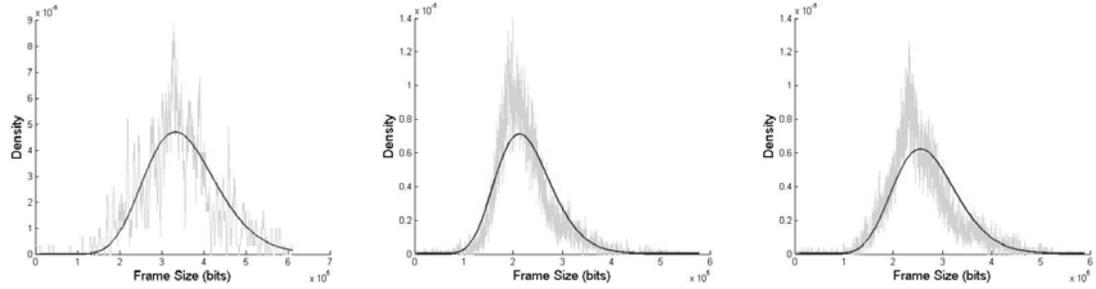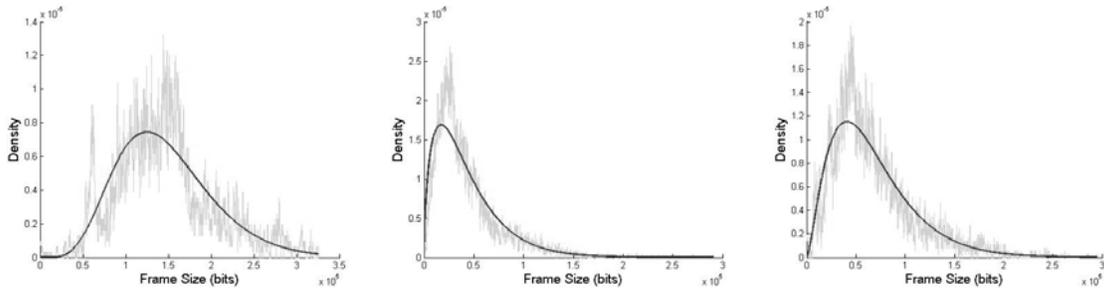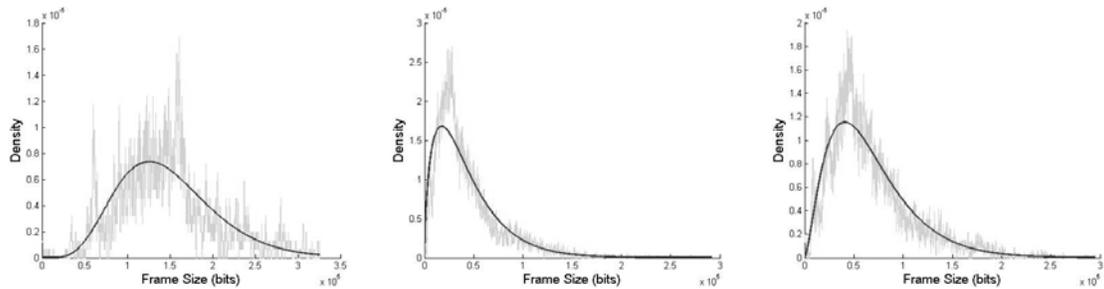
(a) 10-10-10-12 (I/B/P)
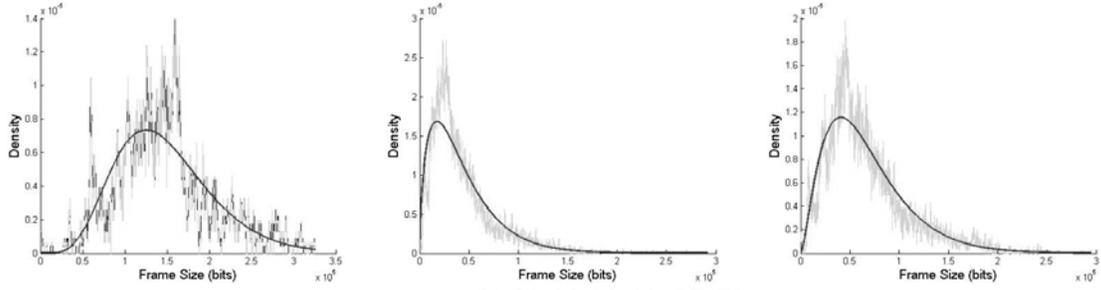
(b) 20-20-20-12 (I/B/P)

(c) 30-30-30-12 (I/B/P)

(d) 20-20-20-3 (I/B/P)

(e) 20-20-20-6 (I/B/P)

Figure 5. Frame size histograms and Gamma models for the various quantization scales

Observing that the derived graphs follow the expected bell-like shape, then the well adopted method of moments is used in order to fit a gamma distribution to the data output. The usual moments approach makes use of the fact that the Gamma distribution has mean $p\mu$ and variance $p\mu^2$. By equating to the mean and sample variance, denoted as m and v respectively, it can be deduced that $\mu=v/m$ and $p=m^2/v$. Therefore, exploiting this relation, the corresponding Gamma-fits to the sample distribution can be derived. Figure 5 illustrates the frame histograms in conjunction with the corresponding Gamma models and Table 2 contains the Gamma distribution parameters for each quantization scale.

Quantization Settings / Frame Types	I Frames		B Frames		P Frames	
	p	μ	p	μ	p	μ
(10,10,10)-12	16.487	21499	15.584	14608	17.071	15895
(20,20,20)-12	6.468	22905	1.682	25572	2.549	26287
(30,30,30)-12	4.406	12235	0.813	7857	1.376	11874
(20,20,20)-3	6.597	22316	1.661	25808	2.542	26446
(20,20,20)-6	6.545	22683	1.676	25720	2.540	26404

Table 2: Gamma model statistics overview of the encoded signals

As a next step, the autocorrelation function is derived for each frame type. Three representative graphs for the case of quantization scale 20-20-20 and Intra-frame period equal to 12, appear in figure 6, suggesting that the autocorrelation exhibits a reduced decay rate beyond the initial lags.
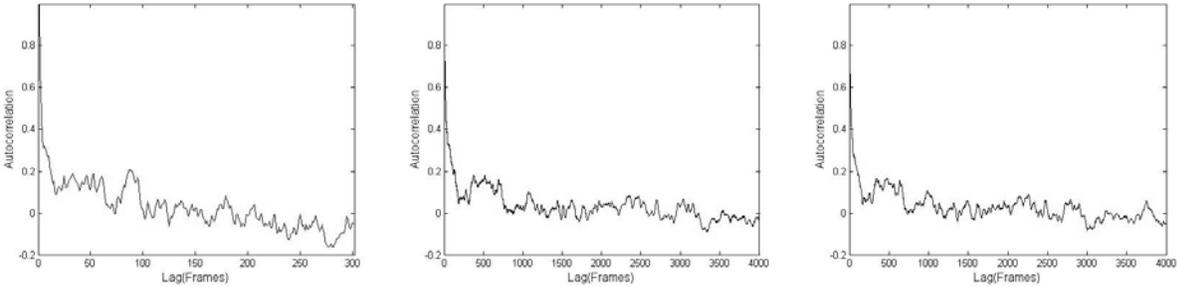
Figure 6. Autocorrelation Graphs for the case of 20-20-20-12 I/B/P frames

It is observed that in the case of H.264 streams, the autocorrelation functions follow the same decaying shape as in previously studied encoding formats, i.e. H.261 (Skianis *et al*, 2003) and MPEG-1(Doulamis *et al,* 2000).

4   Conclusions

This paper reports on an experimental study of H.264 encoded video streams where additional statistical analysis established general results about the video traffic. The experiments covered cases with different quantization scales and GOP lengths, showing that the derived data can be expressed as superimposition of three discrete frame contributions. In this respect, the density functions of the I/B/P frame sizes were derived and it was shown that they can be successfully

represented by Gamma distributions. Moreover, I/B/P autocorrelation functions were drawn, showing that they exhibit an exponentially decaying shape.

Acknowledgement

5 References

Chin H.S., Goodge J.W., Griffiths R. and Parish D.J. (1989), "Statistics of video signals for viewphone-type pictures'', *IEEE Journal on Selected Areas in Communications*, Vol.7, No.5, pp826–832.

Cohen D.M. and Heyman D.P. (1993), ''Performance modeling of video teleconferencing in ATM networks'', *IEEE Transactions on Circuits Systems Video Technology*, Vol.3, No.6, pp408–422.

Doulamis N.D., Doulamis A.D., Konstantoulakis G.E. and Stassinopoulos G.I. (2000), ''Efficient Modeling of VBR MPEG-1 Coded Video Sources'', *IEEE Transactions on Circuits Systems Video Technology* Vol.10, No.1, pp93–112.

Haskell B.G. (1972), ''Buffer and channel sharing by several interframe picturephone coders'', *Bell Systems Technical Journal*, Vol.51, No.1, pp261–289.

Heyman D.P., Tabatabai A. and Lakshman T.V. (1992), ''Statistical analysis and simulation study of video teleconference traffic in ATM networks'', *IEEE Transactions on Circuits Systems Video Technology*, Vol.2, No.1, pp49–59.

Lucantoni D.M., Neuts M.F. and Reibman A.R. (1994),''Methods for performance evaluation of VBR video traffic models'', *IEEE/ACM Transactions on Networking*, Vol.2, No.2, pp176–180.

Maglaris B., Anastassiou D., Sen P., Karlsson G. and Robbins J.D. (1988), ''Performance models of statistical multiplexing in packet video communications'', *IEEE Transactions on Commununications*, Vol.36, No.7, pp834–843.

Nomura M., Fujii T. and Ohta N. (1989), ''Basic characteristics of variable rate video coding in ATM environment'', *IEEE Journal on Selected Areas in Communications*, Vol.7, No.5, pp752–760.

Ramamurthy G. and Sengupta B. (1990),''Modeling and analysis of a variable bit rate video multiplexer'', in: *Proc. of the 7th Internat. Teletraffic Congress Seminar*, Morristown, NJ.

Rodriguez-Dagnino R.M., Khansari M.R.K. and Leon-Garcia A. (1991), ''Prediction of bit rate sequences of encoded video signals'', *IEEE Journal on Selected Areas in Communications*, Vol.9, No.3, pp305–314.

Skianis C, Kontovasilis K, Drigas A. and Moatsos M. (2003), ''Measurement and Statistical Analysis of Asymmetric Multipoint Videoconference Traffic in IP Networks'', *Kluwer, Telecommunication Systems*, Vol.23, pp95-122.

Verbiest W., Pinnoo L. and Voeten B. (1988), ''The impact of the ATM concept on video coding'', IEEE Journal on Selected Areas in Communications, Vol.6, No.9, pp1623–1632.

Wiegand T., Sullivan G., Bjontegaard G. and Luthra A. (2003), ''Overview of the H.264/AVC Video Coding Standard'', *IEEE Tansactions on Circuits and Systems for Video Technology*, Special Issue in H.264.