

PREDICTION METHODS FOR MPEG-4 AND H.264 VIDEO TRANSMISSION

Filip Pilka — Miloš Oravec *

Video services became a large part of internet network traffic. Therefore understanding of video coding standards and video traffic sources, such as video trace files is highly important. In this paper we concentrate on the basic characteristics of mpeg-4 and h.264 video coding standards. We describe the concept of the *i*, *p* and *b* frames in these standards since they are the main feature of every video trace file. Then we describe the content of the video trace files since the trace files are important for researchers to investigate network performance and understanding of network features. These are the important issues in terms of assuring the quality of service (QoS) in multimedia applications spread across the internet. Traffic prediction and bandwidth allocation are the crucial parts in terms of QoS. In this kind of applications, the artificial neural networks are vastly used. Therefore we illustrate the results of neural networks for video traffic prediction using both mpeg-4 and h.264 trace files.

Key words: video encoding, coding standards, prediction, neural networks

1 INTRODUCTION

Multimedia services and applications became the major part of the internet network traffic. Typically video sources are encoded for transport over the internet. The encoded and compressed video sources are characterized by highly variable bit rate, long range dependencies and bursty characteristics. The different types of frames produced by the encoder are responsible for the bursty character of video traffic. The different video frames vary in time and each one contains different amount of information and therefore has different size.

The characteristics of video traffic are contrary to the requirements of the Quality of Service (QoS) of video traffic. Therefore the transport of video over the network is a challenging problem. Some sort of traffic and congestion control needs to be used due to the characteristics of the video traffic [1]. Some of the important procedures that fall within the group of congestion and traffic control are the connection admission control, usage parameter control, traffic shaping, congestion indication, packet discarding *etc.*

One of the schemes for bandwidth allocation is the prediction scheme. The design of bandwidth allocation for video traffic is a complex problem [2]. The artificial neural networks are commonly used for video traffic prediction [3–7] due to their main advantages. They are massively parallel, achieve high computation rates, employ adaptive topology and weights, are capable of modeling complex non-linear mappings *etc.* In this paper we firstly introduce the concepts of video coding using the MPEG-4 and H.264 video standards. Then we introduce the video trace files and a short insight of how video is transmitted through the network. At the end of the paper we present the results of video prediction using artificial neural networks.

1.1 Basic principles of compression algorithms

Every signal, which carries information contains definite amount of redundancy. In our terms every video sequence contains redundancy. The aim of the video compression is to remove the redundancy from the sequence. We recognize two types of compression. The first type is the lossless compression, where only the statistical redundancy is removed from the signal. The outcome of this compression is that on the receiver side, when we want to reconstruct the signal, we get identical copy of the original. This is also a great disadvantage of the lossless compression, because with current technology we can achieve only modest amount of compression of video signals. The second type of compression is the lossy compression with which we can achieve greater compression with the penalty that the reconstructed signal is not identical to the original. The aim of the lossy compression is to meet a given bit rate for storage and transmission of the signal [8].

The goal of every video source coding is the bit rate reduction for storage and transmission. The compression is then done by removing different types of redundancy, spatial, temporal or frequency, which is done at the expense of information loss and by encoding the set of information using an entropy coding technique, such as Huffman coding or Arithmetic coding [9].

2 MPEG VIDEO STANDARD

2.1 Overview of MPEG

The early work on standardization of MPEG audio and video format began in early 1980s. The first efforts to develop unified audio and video standards were within

* Slovak University of Technology in Bratislava, Faculty of Electrical Engineering and Information Technology, Department of Applied Informatics and Information Technology, Slovakia, milos.oravec@stuba.sk

Table 1. MPEG work in progress standards

Name	Description
MPEG-A	Multimedia Application Format
MPEG-B	MPEG Systems Technologies
MPEG-C	MPEG Video Technologies
MPEG-D	MPEG Audio Technologies
MPEG-E	MPEG Multimedia Middleware
MPEG-V	Media Context and Control
MPEG-M	MPEG Extensible Middleware
MPEG-U	MPEG Rich Media User Interface

several international organizations. They were the International Telegraph and Telephone Consultative Committee (CCITT), International Radio Consultative Committee (CCIR) and International Organization of Standardization/International Electrotechnical Commission (ISO/IEC). The MPEG group was established in 1988 within the framework of Joint ISO/IEC Technical Committee on Information Technology with the aim to develop standards for coding of motion pictures, associated audio and the combination of both. The combinations were used for storage and retrieval of digital storage media with bit rates up to 1.5 Mbit/s. The first MPEG standard was issued in 1992 with the name MPEG-1. The later work of MPEG group was focused on updating the original MPEG-1 standard. The MPEG-2 standard was extended to provide better compression algorithms, better suited for applications with higher bit rates than the original MPEG-1. The aim of MPEG-2 was to provide video quality better than NTSC/PAL and up to CCIR 601 with bit rates from 2 to 10 Mbit/s. In the times of the development of MPEG-2 standard there were many new applications emerging. Among others they were the digital cable TV distribution, network services via Asynchronous Transfer Mode (ATM), satellite and terrestrial digital broadcasting. All these services thought to benefit from the increased quality of the new standard. The MPEG-2 was finally issued in 1994 [8]. The standard for scalable and multi-resolution compression should have been contained in the MPEG-3 video standard, which was planned for HDTV compression. This work was finished in the same time as the MPEG-2, so this technology was included into the MPEG-2 standard and the MPEG-3 standard was never issued [10]. The newest standard from the MPEG family is the MPEG-4 visual. The main goal of this standard is audio-visual coding. The main focus of the standard was to extend and update the former MPEG-1 and MPEG-2 standards, so that the new standard would be more flexible and efficient. In 1994 the new H.263 standard was developed with new compression technology. The MPEG group changed its focus from extending the compression algorithms to object coding, which they included into the new MPEG-4 standard. The MPEG-4 visual was finalized in 1998 [9].

The MPEG-7 standard is the fourth approved MPEG standard. The motivation to develop the MPEG-7 stan-

dard was based on the fact, that the potential of the digital video has given to the producers the editing tools to manipulate all the film data faster and with better accuracy. Therefore the standardization of features became more and more important [11]. The MPEG-7 standard is different from the older standards, which represent the content, while the MPEG-7 standard represents information about the content. For this purpose, the standard defines syntax and semantics for describing multimedia contents. The purpose of the standard is to increase the interoperability between different applications and devices that work with audio-visual content descriptions [11].

The MPEG-21 standard was the next step in the MPEG standardization chain. The aim of the standard is to define the technology for the exchange, access, trade *etc* of digital items (video collections, music albums *etc*). Table 1 presents the working standards of the MPEG group [12].

2.2 MPEG-4

Compared to all other “frame-based” video coding standards, the MPEG-4 standard is object based. The model is based on the hybrid DPCM/DCT coding model. Functions of the MPEG-4 model are extended by tools such as enhanced compression efficiency, reliable transmission, coding of separate shapes (objects) in visual scenes, mesh-based compression and human face and body animations. A single application will not need all these tools to be implemented, that is why the standard is divided into series of profiles.

The main features of the MPEG-4 video standard are:

- Compression tools based on ITU-T H.263, which outperforms the compression methods used in MPEG-1 and MPEG-2
- Coding of video objects, which allows separate coding of the foreground and background objects
- Error resilience tools, which assist the decoder to recover from errors during transmission
- Coding of still textures, *ie* also still pictures can be coded by MPEG standard
- Coding of animated visual objects
- Special applications such as “studio” quality video where quality is more important than compression

In order to facilitate the needs of different applications in the MPEG-4 standard, different combinations of tools, objects and profiles were defined. By the definition, a tool is a set of coding functions, which favors a unique feature. An object is a video element that is coded using several tools. And finally, a profile is a set of object types, that the CODEC is able to handle.

In opposition of the traditional view of the video scene being a set of rectangular frames, the MPEG-4 looks on the video sequence as a collection of Video Objects (VO). The video object is an entity that the end user is able to manipulate. It is an area, which can be of any shape and may exist for an arbitrary length of time.

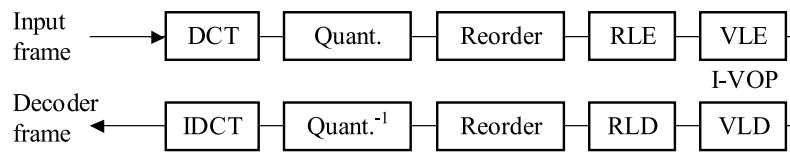


Fig. 1. Encoding and decoding scheme for I-VOP

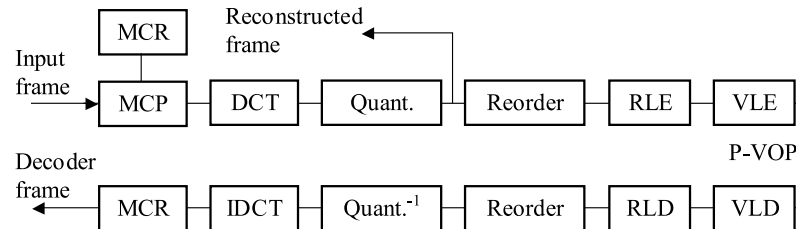


Fig. 2. Encoding and decoding scheme for P-VOP

The instance in which the video object occurs is called the Video Object Plane (VOP). In the traditional view, we can say that the VOP is a single frame and a set of frames forms a VO. However with the new approach we can say that the VOP may be irregularly shaped and occupies only a part of a frame. The separate objects are then coded independently with different resolutions and qualities. Despite the advantages of object based coding, the most popular coding scheme in the MPEG-4 standard is to encode complete frames of video. The tools for this coding scheme are grouped together in the Simple profiles of MPEG-4.

There are three Simple profiles in the standard with the following tools:

- **Simple Profile**
 - I-VOP (Intra-coded rectangular VOP)
 - P-VOP (inter-coded rectangular VOP)
 - Short header (compatibility with H.263)
 - Compression efficiency tools (Four motion vectors per macroblock, Unrestricted motion vectors, Intra prediction, Transmission efficiency tools)
- **Advanced Simple Profile**
 - B-VOP (bidirectionally predicted Inter-coded VOP)
 - Quarter pixel motion compensation
 - Global motion compensation
 - Alternate quantizer
 - Interlace (coding of interlaced video sequences)
- **Advanced Real Time Simple Profile**
 - Improve error resilience
 - Coding flexibility
 - NEWPRED (multiple prediction references)
 - Dynamic resolution conversion

In the next section, we will describe some of the mentioned tools. In Figure 1 and Fig. 2, the encoding and decoding schemes are presented.

The DCT and IDCT blocks represent the forward and inverse Discrete Cosine Transform, which is performed on

the 8×8 blocks of luma and chroma samples. The Quantization block rescales the coefficients on the decoder side. There are two methods described in the standard [12]. The reorder block means that the DCT coefficients are reordered in the zig-zag scan. The RLE block reorders the coefficients once again to represent the zero coefficients efficiently using the Last-Run-Level coding. And finally, the sequence is coded using an Entropy coder. This is the encoding scheme for coding I-VOP.

The P-VOP is coded in the same way but before the application of DCT the motion estimation and compensation [9] needs to be performed. The best method used in every application is left for the designer to choose. The matching region is subtracted from the current macroblock and a residual macroblock is created. Then the same steps as those by the I-VOP are performed. For further motion compensation in the encoder, the quantized residual is rescaled, inverse transformed and stored as a local copy. In the **Advanced Simple Profile**, the B-VOPs are introduced. These use the bidirectional prediction for the motion compensation. Each block or macroblock can be predicted using the forward prediction, backward prediction or an average of the forward and backward prediction. In the forward prediction only a single motion vector, referring to the previous I or P-VOP, is transmitted. In the backward prediction again only one motion vector is transmitted, but this refers to the future I or P-VOP. The bidirectional prediction can be divided into two different types. The bidirectional interpolated prediction transmits two motion vectors, referring to the future and previous VOP. The motion compensation is then produced as an interpolation between the luma and chroma samples in the two reference regions. The second type is the bidirectional direct prediction, where the motion vectors are derived automatically from the motion vectors of the same macroblock in the future VOP. A delta motion vector is then transmitted to correct the automatically created motion vectors. The so called NEWPRED tool is contained in the *Advanced Real Time Sim-*

ple profile, which allows the encoder to select a reference VOP from any of the previously encoded VOPs. This tool allows the decoder to send a feedback message to the encoder, if an error is detected, which identifies the error packet. The encoder chooses a reference VOP foregoing the error packet.

The following list presents the rest of defined profiles within the MPEG-4 standard:

- **The Core profile**
 - B-VOP
 - Alternate quantiser
 - Object-based coding (with Binary Shape)
 - P-VOP temporal scalability
- **The Main Profile**
 - Interlace
 - Object-based coding with grey (“alpha plane”) shape
 - Sprite coding
- **Advanced Coding Efficiency Profile**
 - Quarter pixel motion compensation
 - Global motion compensation
 - Interlace
 - Grey shape coding
 - Shape adaptive DCT
- **The N-bit Profile**
 - N-bit tool – Coding of luminance and chrominance between 4–12 bits per sample (usual 8 bits per sample)
- **Simple Scalable Profile**
 - I-VOP, P-VOP
 - Four motion vectors, Unrestricted motion vectors, Intra prediction
 - Video Packets, Data partitioning, Reversible VLCs
 - B-VOP (Rectangular temporal or spatial scalability 1 enhancement layer)
- **Core Scalable Profile**
 - Includes Simple Scalable Profile tools
 - Rectangular temporal scalability
 - Rectangular spatial scalability
 - Object-based spatial scalability
- **Fine Granular Scalability Profile**
 - B-VOP, Interlace, Alternate quantizer
 - FGS spatial scalability
 - FGS temporal scalability

For texture coding, the MPEG-4 standard has the Scalable Texture Profile and the Advanced Scalable Texture Profile. The Simple Studio Profile and the Core Studio Profile are used for coding of studio quality video.

A more detailed description of all the other profiles in the MPEG-4 visual standard are out of scope of this publication and can be found in [9].

3 H.264/MPEG PART 10

The new video coding standard, entitled the ‘Advanced Video Coding’ is jointly published as Part 10 of the MPEG-4 video coding standard and as ITU-T Recommendation H.264. It was developed by the Moving Picture Experts Group (MPEG) together with the Video Coding Experts Group (VCEG) to outperform the earlier MPEG-4 and H.263 standards.

The encoder and decoder of the H.264 standard use most of the basic functionalities as the previous standards with the exception of the deblocking filter. Important changes occur in the details of each function blocks of the encoder and decoder. Fig. 3 presents the basic encoder of H.264 standard.

Same as in the MPEG-4 standard, each input frame is processed in macroblocks. Each macroblock can be encoded in intra or inter mode. A prediction is made for every macroblock from reconstructed samples. In the intra mode the prediction is done from the current previously encoded slice. In the inter mode the prediction is computed from the motion compensated prediction from one or two reference pictures, which are organized into one or two lists. The prediction reference can be chosen from multiple past or future pictures. The prediction is then subtracted from the current block and a residual block is produced. This block is then transformed, quantized, reordered and entropy coded. The encoded coefficients together with several side information (prediction mode, quantizer parameters, motion vector, *etc*) are passed to the Network Abstraction Layer for transmission or storage. In the encoder, each block is also decoded and reconstructed to provide a reference for future predictions. A filter is applied to reduce the effects of blocking distortions and the reference picture is reconstructed from a series of blocks.

The H.264 consists of three different profiles, where each one includes a set of coding functions:

- **Baseline Profile**
 - Intra and Inter coding (I and P-slices)
 - Entropy coding with Context-Adaptive variable-length codes
- **The Main Profile**
 - Interlaced video
 - Inter coding (using B-slices)
 - Inter coding using weighted prediction
 - Entropy coding with Context-Based arithmetic coding
- **Extended Profile**
 - Efficient switching between bitstreams (SP and SI-slices)

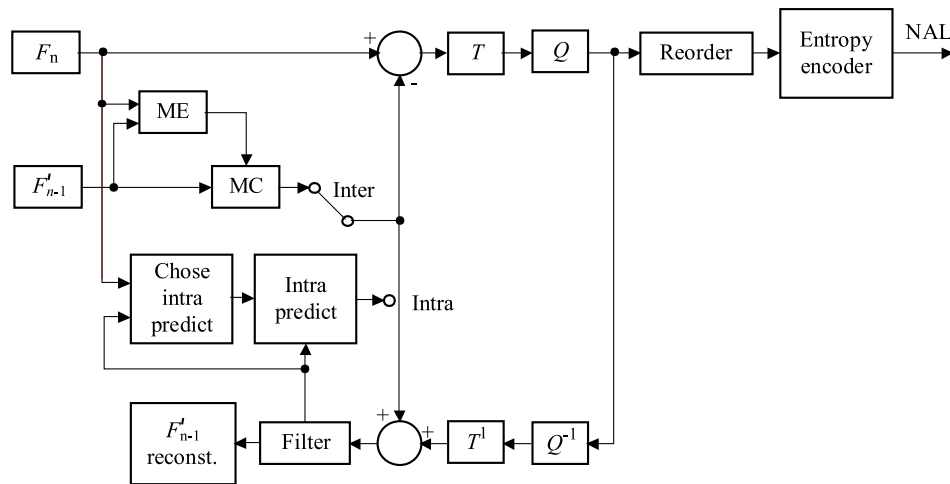


Fig. 3. The encoder of H.264

o Error resilience

In the standard, the Video Coding Layer and the Network Adaptation Layer are defined. The Video Coding Layer represents the encoding of video data and the output is then called the VCL data. The VCL data is then mapped to the NAL units. The purpose of this separation is to distinguish the coding and the transport specific features.

The H.264 encoder may use one or two reference pictures from a number of previously encoded pictures for motion compensation. This allows the encoder to search for the best match, what is better approach than just using the previously encoded picture. The encoder and decoder thus store one or two lists containing the reference pictures. The inter coded macroblocks (P-slices) are predicted from the pictures in a single list, the so called list 0. Inter coded macroblocks and macroblocks of B-slices may be predicted from two lists, list 0 and list 1.

A slice is a sequence of macroblocks grouped together. The number of macroblocks in a slice does not need to be constant. There are five types of slices in the H.264 standard and each picture may be encoded as a mixture of different slice types. As stated before, **the Baseline Profile** supports I and P-slices. In I-slices, the intra coded macroblock regions, 16×16 or 4×4 luma and 8×8 chroma are predicted from the previously coded samples in the same slice. P-slices may contain intra, inter coded or skipped macroblocks. The inter coded macroblocks are predicted using motion compensation with quarter sample accuracy. After the prediction the data for each macroblock are transformed using a 4×4 integer transform (DCT based transform), quantized, reordered and entropy coded using a context-based variable length coding. All other elements are coded using fixed length or Exponential-Golomb variable length codes.

The difference in the block-based motion compensation in the H.264 standard, compared to the other standards, is the support for a range of block sizes and a sub-

sample motion vectors (quarter-sample resolution in the luma components). The luminance components may be split up in four ways:

- One 16×16 macroblock partition
- Two 16×8 partitions
- Two 8×16 partitions
- Four 8×8 partitions If the 8×8 partitions are chosen, the blocks may be split up further:
- One 8×8 sub-macroblock partition
- Two 8×4 sub-macroblock partitions
- Two 4×8 sub-macroblock partitions
- Four 4×4 sub-macroblock partitions

This method of partitioning macroblocks is known as Tree Structured Motion Compensation. For each partition, a separate motion vector is needed and each motion vector needs to be encoded and transmitted.

As stated before, a filter is applied to each decoded macroblock. It is the deblocking filter, which reduces the blocking distortions. The filter is applied after the inverse transform in the encoder. The filter smooths the block edges and the filtered image is used for motion compensated prediction. This can improve the compression performance since the filtered picture is a better reproduction of the original frame.

The extension of **the Main Profile** consists of tools supporting the B-slices, weighted prediction, support for interlaced video and alternative entropy coding based on Arithmetic Coding. The B-slices work similarly as in the MPEG-4 standard – that means they use the motion compensated prediction. The B-slices can be predicted from one or two reference pictures, before or after the current picture. The reference pictures are stored in two lists. They contain both short term and long term pictures.

The prediction of macroblock partitions can be done in one of four ways. They are the Direct prediction, Motion compensated prediction from list 0, Motion compen-

sated prediction from list 1 and Motion compensated bi-predictive prediction from list 0 and list 1.

In the **Extended Profile**, the SP and SI-slices were introduced. They are specially designed to enable efficient switching between video streams and for efficient random access. The SP-slices support switching between similar coded sequences without increased bitrate. The SP-slices are coded using motion compensated prediction and at the switching point there are three SP-slices

The additional switching slice in the Extended Profile is the SI-slice. It is used in similar way than the SP-slice, except the prediction is made up of the 4×4 Intra prediction modes. These slices are used when switching two completely different sequences.

More detailed description of the H.264 standard can be found in [9].

4 TRACE FILES

As described above, the MPEG standard introduced the I, P and B frames, similar to the frame types in H.264 video standard. These types of frames are grouped together in a specific order, which are called the groups of pictures (GoP). One GoP is defined as a sequence of frames beginning with the I frame until the next I frame. A typical GoP for the MPEG standard includes three P frames and two B frames before and after each P frame.

The above described video standards are only a part of the variety of encoding schemes used for video coding [8], [13]. Video traces are typically used in studies on transporting video through any kind of network. The simplest of these trace files is the so called video frame size trace [13]. This kind of traces contains the sizes of individual frames. There are also more complex video traces, which contain numerous frame qualities. The qualities given by the trace files can be:

- Number of video frame
- Frame period
- The cumulative display time
- Frame size
- Frame type
- Quality of luminance component
- Quality of the chrominance components.

The trace files are then typically an ASCII file with one line per frame. Traces that contain only the frame sizes are called terse trace files. The trace files containing several of the qualities described above are called verbose trace files. Some traces give the frames in the display order and some in the encoded order. The frame index however always represents the display order. The encoded order means that the P frames are given before the B frames, because, they are needed to decode the successive B frames [13].

5 TRANSMISSION OF VIDEO OVER THE INTERNET

The main protocols for transmission of video over the Internet are the Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP) [14]. The TCP was not designed or optimized for real time applications such as video streaming. The congestion window in the TCP protocol is unceasingly changing due to the slow start algorithm. The consequences of this algorithm are fluctuations in the video transmission. The retransmission mechanisms in the TCP can be a problem as well in video streaming applications. The common idea of the retransmission is, that based on the real time nature of video streams all re-send data will arrive after they had to be displayed. The packet retransmission also causes high levels of end to end latency [15]. On the other hand, the UDP may seem ideal for video transmission over the Internet. It has no congestion control, nor retransmission mechanisms. That means that the transmission rate of the UDP based protocols will be constant and in case of congestion the TCP based transmissions will activate their congestion mechanisms and all the bandwidth will be available to the UDP streams. Notwithstanding the mentioned advantages of the UDP-based protocols, also this scheme has its drawbacks. The first is that the lack of congestion control may lead to total collapse of the network. And secondly, when transmission is done in a noisy environment the packet loss in UDP-based protocol is high. With no retransmission mechanisms the quality of the video will be low [14]. The lack of retransmission mechanisms nowadays is partly eliminated with the error resilience tools included in the video coding standards, such as MPEG-4.

Based on the drawbacks of both of the main Internet protocols researchers still develop new methods for video transmission [14],[16-18].

6 NEURAL NETWORKS

In this paper the neural networks are used as a tool for video traffic prediction. Used artificial neural network [19] consisted of standard parts (input connections also known as synaptic weights, sum function to compute the weighted sum of the inputs, nonlinear activation function and a bias (or a threshold))

Neurons are then arranged to different types of topologies. The so called multilayer perceptron (MLP), which has fully connected layers of neurons, is mostly used. The main idea of neural networks is to adapt (train) the weights of the neurons to get the desired outputs [20]. In the MLP neural network, the adaptation is done using the backpropagation algorithm or one of its modifications. The computational ability, nonlinear characteristics and accuracy of neural networks are the main skills that place neural networks on top of the tools for predictions [20]. In this paper we present the prediction results of the MLP

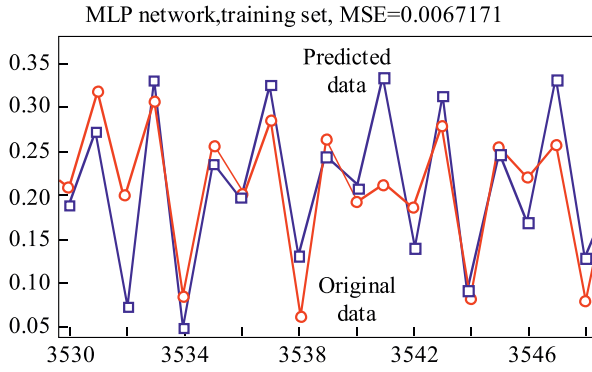


Fig. 4. Training set for MPEG trace file

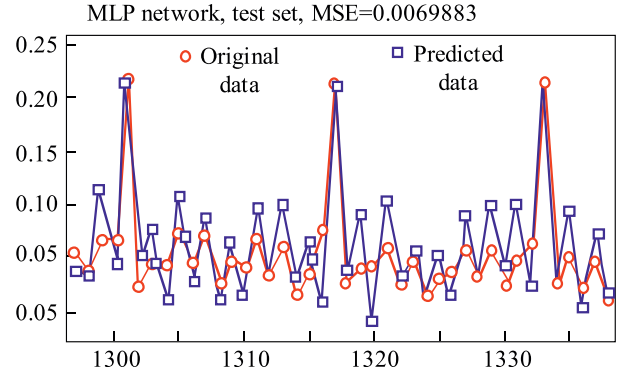


Fig. 5. Test set for MPEG trace file

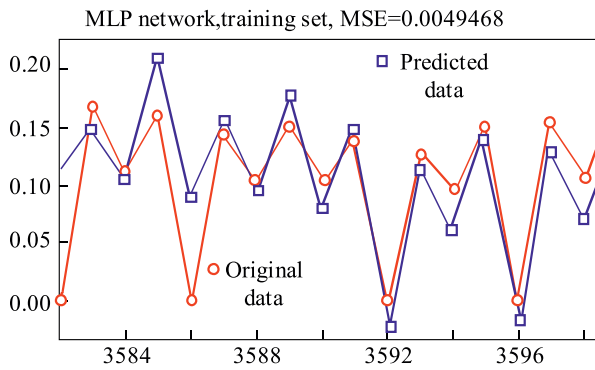


Fig. 6. Training set for H.264 trace files

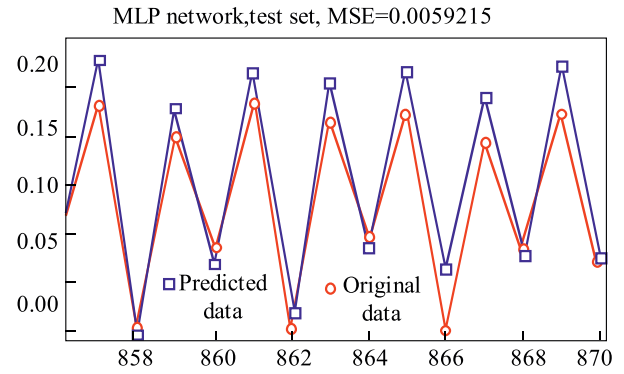


Fig. 7. Test set for H.264

neural networks that were trained with the Levenberg-Marquardt learning algorithm [21].

7 SIMULATIONS

For the simulations, we have used the trace files from the movie *Silence of the Lambs* [13]. We have used the MPEG-4 and H.264 traces for our simulations to make a comparison which of the two formats is better suited for frame predictions. We have taken the first ten thousand frames as our data set. The data set was divided into five separate sets and each set was taken as training set and the other four sets were taken as testing sets (5-fold validation). So for each trace file five simulations were done. The values of MSE for each set, the mean value and the variance of the errors were taken as results. A simple MLP with twelve input units, twenty five hidden neurons and one output neuron was used for simulations. Logistic function was used as the activation function in the hidden neurons and linear activation function was used in the output neuron. For illustration, we present results for only one training/testing set combination for each video format. Figures 4 and 5 present the detail of the prediction by the neural network for the MPEG-4 trace file.

Figures 6 and 7 present detailed results of the neural network for the H.264 trace file. In Tabs. 2 and 3, we present the actual values of MSE (mean-square error) of

each set with mean value and the variance of the errors for MPEG-4 and H.264 traces.

8 CONCLUSION

In this paper we presented the fundamentals of video compression algorithms used in the MPEG and H.264 video coding standards. We presented the tools used in different profiles described by the standards, which are important in terms of understanding the evaluating the video trace files used for prediction of video streams for network transmission.

In the last section, we presented the results of prediction simulations using artificial neural networks. We presented the results of the multilayer perceptron for MPEG-4 and H.264 video trace files. From the MSE values presented in Tabs. 2 and 3 and the variance of the errors, we can see that the neural networks are suitable to predict video streams. The differences between the coding schemes are small and therefore we assume that both the video coding standards can be predicted using artificial neural networks.

Compression methods used in the video coding standards and prediction methods are the key to achieve required QoS for nowadays multimedia traffic in the internet. The compression methods assure that only relevant information is transmitted and the prediction schemes care for efficient bandwidth allocation.

In our future work we want to focus on application of recurrent neural networks for traffic prediction and the capability of recurrent neural networks to perform multi-step ahead prediction.

Table 2. Error values for MPEG files

Sets	Training	Test
Set 1	0.00383	0.01851
Set 2	0.00574	0.00482
Set 3	0.00672	0.00699
Set 4	0.00714	0.00281
Set 5	0.00617	0.00255
Mean	0.00592	0.00714
Variance	1.64607×10^{-6}	4.36353×10^{-5}

Table 3. Error values for H.264 files

Sets	Training	Test
Set 1	0.00358	0.01235
Set 2	0.00465	0.00524
Set 3	0.00495	0.00592
Set 4	0.00622	0.00236
Set 5	0.00550	0.00278
Mean	0.00498	0.00573
Variance	9.73512×10^{-7}	1.60421×10^{-5}

Acknowledgement

Research described in this paper was done within the grant No. 1/0214/10 of the Slovak Grant Agency VEGA.

REFERENCES

- [1] ABDENNOUR, A.: Short-term MPEG-4 video traffic prediction using ANFIS, *International Journal of Network Management* **6** No. 15 (2005), 377–392.
- [2] GUOQIANG, M.—HUABING, L.: Real Time Variable Bit Rate Video Traffic Prediction, *Int. Journal of Communication Systems* **4** No. 20 (2007), 491–505.
- [3] ESWARADASS, A.—SUN, X.—WU, M.: A Neural Network Based Predictive Mechanism for Available Bandwidth, *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05)* **1** No. 1 (April 2005).
- [4] POPESCU, I.—CONSTANTINOU, P.—NAFORNITAM.—NAFORNITA, I.: Generalized regression neural network prediction model for indoor environment, *IEEE Symposium on Computers and Communications* No. 2 (2004), 657–661.
- [5] FRANK, R. J.—DAVEY, N.—ANDHUNT, S. P.: Time Series Prediction and Neural Networks, *Journal of Intelligent and Robotic Systems* **1-3** No. 31 (2001), 91–103.
- [6] BECERIKLI, Y.—OYSAL, Y.: Modeling and prediction with a class of time delay dynamic neural networks, *Applied Soft Computing* **4** No. 7 (2007), 1164–1169.
- [7] ORAVEC, M.—PETRÁŠ, M.—PILKA, F.: Video Traffic Prediction Using Neural Networks, *Acta Polytechnica Hungarica* **4** No. 5 (2008), 59–78.
- [8] SIKORA, T.: MPEG digital video-coding standards, *Signal Processing Magazine IEEE* **5** No. 14 (1997), 82–100.
- [9] RICHARDSON, I. E. G.: H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia, John Wiley & Sons, Ltd. (2003).
- [10] WATKINSON, J.: MPEG Handbook, Newton, MA, USA: Butterworth-Heinemann (2001).
- [11] ČERNEKOVÁ, Z.: Temporal video segmentation and video summarization, Dissertation Thesis, Bratislava: FMPI Comenius University (2009).
- [12] GROUPMPEG. MPEGSTANDARDS[ONLINE]: <http://mpeg.chiariglione.org/standards.htm>.
- [13] SEELING, P.—REISSLEIN, M.—KULAPALA, B.: Network Performance Evaluation with Frame Size and Quality Traces of Single-Layer and Two-Layer Video: A Tutorial, *IEEE Communications Surveys and Tutorials* No. 6 (2004), 58–78.
- [14] HA, R.—ZHUANG, W.: Split-domain video transmission protocol for video streaming over hybrid wired-wireless connections, *Journal of Visual Communication and Image Representation* **4-5** No. 16 (2005), 450–474.
- [15] KRASIC, B.—LI, K.—WALPOLE, J.: The Case for Streaming Multimedia with TCP, *Lecture Notes In Computer Science, In proceedings of the 8th International Workshop on Interactive Distributed Multimedia Systems* No. 2158 (2001), 213–218.
- [16] BŘEZINA, A.—POLEC, J.—PAVLOVIČOVÁ, J.—KOLINGEROVÁ, I.—BANDZI, P.: Region-Based Texture Coding at Very Low Bitrates, *Journal of Electrical Engineering* **1-2** No. 56 (2005), 3640.
- [17] PAVLOVIČOVÁ, J.—POLEC, J.—KELEŠI, M.—MOKOŠ, M.: Detection of Transmission Errors in Block Coded Images, *Journal of Electrical Engineering* **5-6** No. 56 (2005), 121–127.
- [18] KUO, C.-C.—CHEN, M.-S.—CHEN, J.-C.: An Adaptive Transmission Scheme for Audio and Video Synchronization based on Real-time Transport Protocol, *IEEE International Conference on Multimedia and Expo (ICME'01)* **104** No. 2001.
- [19] BALAGUER, E.—PALOMARES, A.—SORIA, E.—MARTINGUERRERO, J. D.: Predicting service request in support centers based on nonlinear dynamics, ARMA modeling and neural networks, *Expert Systems with Applications* No. 34 (2008), 665–672.
- [20] VLAD, S.: On the Prediction Methods Using Neural Networks, *6th International Symposium of Hungarian Researchers on Computational Intelligence*, Budapest (2005).
- [21] KISI, O.: Multi-layer perceptrons with Levenberg-Marquardt training algorithm for suspended sediment concentration prediction and estimation, *Hydrological Sciences Journal* **6** No. 49 (2004), 1025–1040.

Received 3 June 2010

Filip Pilka received the MS degree in Telecommunications from Slovak University of Technology in Bratislava, Slovakia, in 2008. Since 2008 he has been a PhD student at Department of applied informatics and information technology of the same university. He has worked in area of artificial neural networks, machine learning and prediction.

Miloš Oravec received the MSc, PhD and Assoc Prof degrees in telecommunication engineering and applied informatics in 1990, 1996 and 2002. Now, he is the Associate Professor at the Department of Applied Informatics and Information Technology, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology. He is a member of IET. His research interests include artificial intelligence, neural networks, biometrics, data analysis and prediction, signal processing, and communication networks and services.