FACTA UNIVERSITATIS (NIŠ)
SER.: ELEC. ENERG. vol. 16, Dec. 2003, 401-414

Fractal and Multifractal Analyses of Compressed Video Sequences

Irini Reljin and Branimir Reljin

Abstract: The paper considers compressed video streams from the fractal and multifractal (MF) points of view. Video traces in H.263 and MPEG-4 formats, generated at the Technical University Berlin and publicly available, were investigated. It was shown that all compressed videos exhibit fractal (long-range dependency) nature and that higher compression ratios provoke more variability of the encoded video stream. This conclusion is approved from the MF spectra of frame size video traces. By analyzing individual frames and their MF spectra the additive nature is approved.

Keywords: Signal processing, fractal behaviour, multifractal spectrum, standard H.236, MPEG-4.

1 Introduction

Images, in general, and digital images as well, are characterized by a large amount of information. Consider, for instance, still images. A standard quality monochrome (black-and-white) image comprises spatial resolution of 512×512 pixels with 8bits (1 byte, in short, 1B) pixel depth (ie., 256 magnitude levels). In this way one image occupy 2 Mbits or 260 kB of information. Color image (RGB = red, green, blue) of the same size needs 3 times more data. For medical applications, a digital magnetic resonance (MR) image is characterized by 512×512 pixels, and 12 bpp (bits per pixel).

Manuscript received June 22, 2003. An earlier version of this paper was presented at the 10th Telecommunications Forum TELFOR 2002, November 26-28, 2002, Belgrade, Serbia.

I. Reljin is with the PTT College, Zdravka Čelara 16, 11000 Belgrade, Serbia and Montenegro (e-mail: ireljin@ptt.yu). B. Reljin is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, Serbia and Montenegro (e-mail: reljin@etf.bg.ac.yu).

For practical point of view two bytes per pixel is assumed, so such an image contains 0.5 MB or 4 Mbits of information. Furthermore, an X-ray image (digitized from X-ray film or generated directly from digital radiology equipment) is of the size of 2048×2048 (even 4096×4096) pixels with depth 12-16 bpp, i.e., 8 (32) MB, or 64 (256) Mbits [1]. Such files are not so large but a typical radiology session needs 60-100 MR images or, 2-6 X-ray images for medical examination, i.e., 30-50 MB (240-400 Mbits) of information. Transferring, processing and archiving such kind of information is not so difficult only if a small number of images are used. However, in typical clinical applications, for instance, for a medium sized hospital (300 beds) the amount of daily generated data exceeds 3 GB, or more than 1 TB per year [2].

Difficulties in image handling are dramatically increased when image sequences are used, such as video streams for commercial and/or entertainment purposes, since the minimum frame rate is about 10 frames per second (frame/sec). Even for the low quality video stream: for instance, a monochrome video with 800×600×8 pixels per frame and 10 frame/sec, the bit rate needed is almost 40 Mb/s, i.e., 17 GB of information for one hour of such video. The large volume of image data highly requires the *image compression.*

There are two main criteria by which the image compression can be classified. One is the image quality after compression and the second is the complexity of the compression method. The main goal in all of compression techniques is to achieve as high as possible compression ratio with invisible image degradation, by exploiting as simple as possible compression (decompression) algorithms. The compression methods can be classified in two main groups as *lossless* and *lossy* techniques [1],[3],[4]. A lossless method is one that allows exact reconstruction of all of the individual pixel values, while a lossy method does not. Lossless algorithms eliminate only redundant information and these methods are often referred to as image coding rather than compression. Conversely, lossy compression algorithms eliminate irrelevant information as well, and thus permit only an approximate reconstruction of the original. Certainly, lossy compression algorithms achieve higher compression ratios. Usually, a trade-off between image quality (fidelity of reconstruction) and the compression ratio is performed for a particular application. Note that in medical applications only lossless compression is acceptable in the case when the medical diagnosis is expected, after the examination of medical images. For commercial and entertainment purposes lossy compression is widely used.

A complexity of the compression algorithm determines the time required

for file processing. This is particularly important when images are being compressed for real time transmission, as in videoconferencing, video telephony and interactive video. Note here that the algorithms that achieve the densest compression are not usually the fastest, so the choices have to be made for a particular application [5].

All compression algorithms are based on the high correlation between adjacent pixels in the image frame. By exploiting this feature the difference between neighboring pixels can be used for image coding and, in this way, instead of requiring (for instance) 8 bits per pixel, less bits are sufficient. Further improvements in video compression use the concept of time prediction: the pixel values in next frame are predicted from several preceding pixels and then only the differences from actual and predicted values (which are usually small) are stored and used for processing. Moreover, instead of pixel value the group of pixel values is used in the prediction process. For video streaming the correlation between successive frames as well as the subjective characteristics of the human visual system are exploited leading to the high compression ratios (more than 150:1) permitting multimedia video streaming over commercially available communication systems. Several standards, combining a number of different compression methods, are now defined, adopted, and successfully embedded into the high-speed hardware [1]-[5].

This paper considers the characteristics of the compressed video streams from the fractal and multifractal point of view. We used publicly available video traces generated at the Technical University Berlin [6]. Those video traces have been generated from MPEG-4 and H.263 encoders covering the range from very low bit rates (as for wireless communication) to bit rates and quality levels beyond HDTV (high definition television). In Section 2 the brief review of the video compression methods is exposed. Section 3 considers the fractal and multifractal analyses of several videos, H.263 and MPEG-4 encoded for different compression ratios, permitting different image quality. Several concluding remarks are derived in Section 4.

2 Brief Review of the Video Compression Methods

Video compression standards have been originated since 1984, with the standard H.261, defined by the ITU-T Study Group 15 (*Transmission Systems and Equipment*) for video telephony and video-conferencing applications, emphasizing low bit rates and the low coding delay. This standard was intended for audiovisual low-bit-rate ISDN services. At the beginning, the

design target was for p*384 Kb/s, where p was between 1 and 5. From 1988 the focus shifted at bit rates around $p * 64$ Kb/s, where p is from 1 to 30, whence came the alternative name $p * 64$ for the standard [4]. In fact, $p * 64$ (or H series standards) is a group of audiovisual teleservices standards consisting of H.221 frame structure and multiplexing; H.230 frame synchronous control; H.242 communication between audiovisual terminals (signaling protocol); H.320 systems and terminal equipment; and H.261 video codec (coder and decoder). Audio codecs at several bit rates have also been specified by other ITU-T recommendations known as G standards (G.711, G.722, G.728). Standard H.261 was designed for video telephony and videoconferencing, in which typical scenes are basically static, composed of talking persons (the so-called *head-and-shoulder sequences*), rather than general TV programs that contain a lot of motion and scene changes [4]-[5].

Further improvements in video coding were incorporated in the standard H.263 adopted in 1996. The design target was to obtain video streaming with bit rates lower than 64 Kb/s (known as a *very-low-bit-rate*); for sending video data across the PSTN (*Public Switched Telephone Network*) and the wireless (cell phone) network. During the development of H.263 two different goals were identified: the near-term goal would be to enhance H.261 using the same basic principles, and the long-term goal would be to design a new video- coding standard that may be fundamentally different from H.261. As a result, the near-term effort leads to H.263 and H.263+ (or H.263 Version 2), while the long-term effort is now referred to as H.26L (previously called H.263L) which had been scheduled for adoption in July 2002 [5] while it was renamed to H.264 AVC in Summer 2003.

The coding algorithm used in H.261 is a hybrid of motion compensation to remove temporal redundancy and transform coding to reduce spatial redundancy. Such a framework forms the basic of all video-coding standards that were developed later.

H.261 defines a standard video input called the *Common Intermediate Format* (CIF) assuring the compatibility to standard TV. It uses a sequence of frames and the maximum frame rate is specified to be 30/1.001 (approx. 29.97) frame/sec, which is the same as in NTSC (*National Television System Committee*) TV standard adopted in North America and Japan. In Europe and many other countries, where PAL (*Phase Alternation Line*) TV standard is adopted, the frame rate is 25 frame/sec. The minimum permitted frame rate is a quarter of maximum (7.49 frame/sec). Each frame consists of 288 non-interlaced lines, each having 352 luminance pixels. Color sampling is at half the rate of luminance, both horizontally and vertically; depending on the

realization this sampling structure is performed as 4:1:1, or 4:2:0 - the last one exhibits better performances [4]. The pixel depth is 8 bpp. At maximum frame rate of 29.97 frame/sec (NTSC) or 25 frame/sec (PAL) the input video bit rate is 36.5 Mb/s or 30.5 Mb/s. For reasons of interoperability and low cost, a lower resolution format, called QCIF (Quarter CIF) has also been defined in H.261. It has half the resolution of CIF, both horizontally and vertically, and, consequently, the quarter memory and bit rate requirements.

Although, historically, H.261 started 2 years before JPEG (*Joint Photographic Experts Group*) still image compression standard, an improved version of H.261 ($p*64$ standard), started from 1988, used several benefits from JPEG, such as, for instance, intraframe DCT (*Discrete Cosine Transform*), RLC (*Run Length Coding*) and VLC (*Variable Length Coding*), but introduced also a block-based motion-compensated interframe coding. That is, the picture data in the previous frame can be used to predict the image blocks in the current frame, and, as a result, only differences, typically of small magnitude, between the displaced previous block and the current block have to be transmitted [4]-[5].

As noted earlier, H.26x standards are derived primarily for video telephony and videoconferencing applications, assuming mainly quasi-static images. For entertainment video applications the *Moving Pictures Expert Group* (MPEG), established in 1988 from the ISO (*International Standard Organization*), standardized a coded representation of video and associated audio suitable for digital storage (magnetic disks, solid-state memories, optical CD-ROMs, digital audio tape, etc.) and transmission media. As a result several standards known as MPEG-x (x being the corresponding integer number starting from 1) are derived and adopted [4]-[5].

The MPEG-1, adopted in 1993, has been primarily developed for coding moving pictures or similar audiovisual signals at about 1.5 Mb/s, for storing them on a CD with a quality comparable to VHS (*Video Home System*) cassettes. The straightforward extension of MPEG-1 leads to the MPEG-2 coding scheme, adopted in 1995, which is flexible enough to handle a range of video applications with different bandwidth constraints and picture qualities. The MPEG-2 is downward compatible to MPEG-1 but permits standard TV quality pictures and even HDTV quality. This standard is used not only for Video-CD, DVD (*Digital Versatile* (or, *Video*) *Disk*) but also for digital cable and broadcasting TV applications [4]-[5].

In contrast to the "frame-based" video coding of MPEG-1, MPEG-2 and H.263, the MPEG-4 standard (working draft in 1996, adopted in 1999) is object-based [5]. Each scene is composed of Video Objects (VOs) that are

coded individually. (If scene segmentation is not available or not useful, the standard defines the entire scene as one VO.) Each VO may have several scalability layers (one basic layer and one or several enhancement layers) referred to as Video Object Layers (VOLs). Each VOL in turn consists of an ordered sequence of snapshots in time, referred to as Video Object Planes (VOPs). For each VOP the encoder processes the shape, motion and texture characteristics [5]-[6]. This standard is developed to address the emerging needs of integrating the communications, TV/film/entertainment, and different Web-based services usually known as *multimedia*. Further investigations in standardization of multimedia content description are reported as MPEG-7 and MPEG-21 standards, not yet commercially available. (MPEG-7 is available as a reference software, known as *eXperimentation Model*= XM, while for MPEG-21 the latest document is the MPEG-21 *Multimedia Framework*) [5].

A basic video coding scheme in video compression standards is as follows. Each picture (frame) is divided into a number of blocks, which are grouped into macroblocks. For MPEG-1/2 and H.26x standards blocks are composed of 8×8 pixels (luminance or chrominance). Four luminance blocks plus blocks with chrominance values form a macroblock. The number of chrominance blocks in a macroblock depends on the sampling format: formats 4:2:0, 4:1:1, 4:2:2 and 4:4:4 are used. The first frame in video sequence is encoded in intraframe coding mode (I-frame), exploiting spatial redundancy of the frame pixels, without reference to any past or future frames, similar to JPEG still-image coding. For instance, at the MPEG-1 encoder the DCT is applied to each 8×8 luminance and chrominance block and the RLC and VLC are applied to DCT coefficient. In this way I-frames are low compressed. Each m-th frame (where m depends on the coding scheme) is coded as I-frame. Since I-frames can be decoded without knowing anything about other pictures in video stream, they can serve as random access points to the video material [4]-[6].

Each subsequent frame is coded using interframe prediction (P-frames), which means that only data from the nearest previously coded I- or P-frame is used for prediction. P-frames contain motion compensation and provide more compression than I-frames; for instance, P-frame contains 50 to 70% less number of bits needed for an I-frame. However, coding errors can propagate between P-frames. Since P- frames are usually used as a reference for the prediction for future or past frames, they provide no suitable access points for random access functionality or editability, such as FF/FR (*Fast-Forward* and *Fast-Reverse*) options when searching video material [5].

By introducing new frames, bi-directional predicted frames (B-frames), high compression and reasonable random access and FF/FR functionality is achieved. B-frames are coded using motion-compensated prediction from nearest past- (forward prediction) or future (backward prediction) already coded frames, either I- or P-frames. B-frames require approximately 50% of the number of bits needed for a P- frame (15-25% of bits needed for I-frame). Since B-frames are not used as a reference they do not propagate errors. But, note that the backward prediction is possible if the frames are reordered and transmitted so that the future frame is received before the current B-frame. This reordering process at the coding and decoding stage introduces significant delays depending on the number of the B-frames between two reference frames. Besides the picture reordering, B-frames also require more memory in the decoder [5].

The MPEG algorithms allow the encoder to choose the right combination of frame types, in a repeating sequence, to obtain desired performances: compression ratio, random accessibility and picture quality. This frame type sequence is called a group of pictures (GoP) which are generally specified by two parameters: m, which defines a number of B- and P-frames between two I-frames in the data stream sequence, and n, which defines the number of successive B- frames between two I- and/or P-frames [6]. As a general rule, a video sequence coded using I-frames only (I I I I I I ...) allows the highest degree of random access, FF/FR and editability, but achieves only low compression. A sequence coded with I- and P-frames (I P P P P P I P P ...) achieves moderate compression and a certain degree of random access and FF/FR functionality. A sequence containing all three types of frames (I B B P B B P B B I B B P ...) may achieve high compression and reasonable random access and FF/FR functionality, but significantly increases the coding delay, which may not be tolerable for video telephony or videoconferencing [4]-[6].

Note that H.263 standard uses the PB frames - the frames consisting of two pictures (P- and B-) being coded as one unit, instead of B-frames. With this coding option the picture rate can be increased considerably without substantially increasing the bit rate. For instance, for the same frame rate, H.263 gives 30% better bit rate than MPEG-1.

3 Fractal and Multifractal Performances of Compressed Video

The Telecommunication Networks Group at the Technical University Berlin [6] have generated the library of frame size traces of long MPEG-4 and H.263 encoded videos, since these standards are expected to be used in future wireless networks. They used several hit movies from VHS videotapes and video material from cable TV. The video traces were grabbed at the frame rate of 25 frames/sec in the QCIF format, that is with a luminance resolution of 176×144 pixels and 4:1:1 (Y:U:V) sampling format with a pixel depth of 8 bits.

The uncompressed YUV video information was encoded into an H.263 bit stream and into an MPEG-4 bit stream [6]. H.263 was targeted to four bit rates: 16 Kb/s, 64 Kb/s, 256 Kb/s and variable bit rate (VBR), i.e., without setting a target bit rate. For MPEG-4 three different quality levels were selected: low, medium and high. In [6] statistical analysis of the frame-size video traces is performed. Their analyses show the intuitively expected tendency: the higher compression ratios the more variability of the encoded video stream.

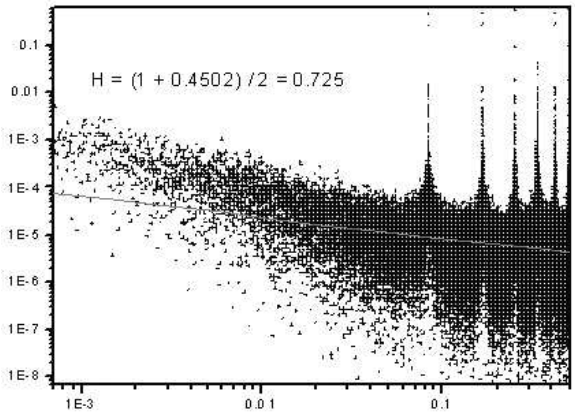

Fig. 1. The periodogram of MPEG-4 low-quality encoded movie "Mr. Bean".

A compressed digital video and modern communication traffic as well, are characterized by burstiness, exhibiting thus long-range dependency [7]-[10]. Our previous work [11]-[13] was concentrated to the fractal and multifractal analysis of MJPEG (Motion JPEG) and MPEG-1 encoded movie "Star Wars" [14]. Here, we used video traces as in [6] and performed fractal

and multifractal analyses over them. The fractal behavior [15]-[16] was investigated through the Hurst index value, determined from RS diagram, the periodogram and IDC (index of dispersion) methods. All video traces exhibit fractal behavior: the Hurst index was between 0.5 and 1.0. As an illustrative example, in Fig. 1 the periodogram, derived for low-quality MPEG-4 encoded movie "Mr. Bean", is depicted.

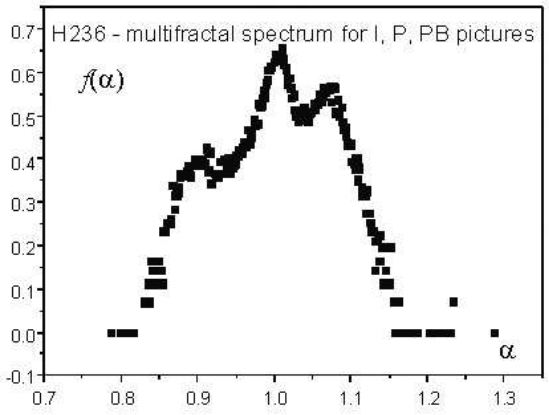

Fig. 2. MF spectrum of the frame size video traces for H.263 16Kb/s encoded movie "Mr. Bean".

All compressed videos are then analyzed from the multifractal (MF) point of view. MF spectra were derived from histogram method [15], by using computer program derived in [11] and [17]. Some characteristic results are depicted in Figs 2-5.

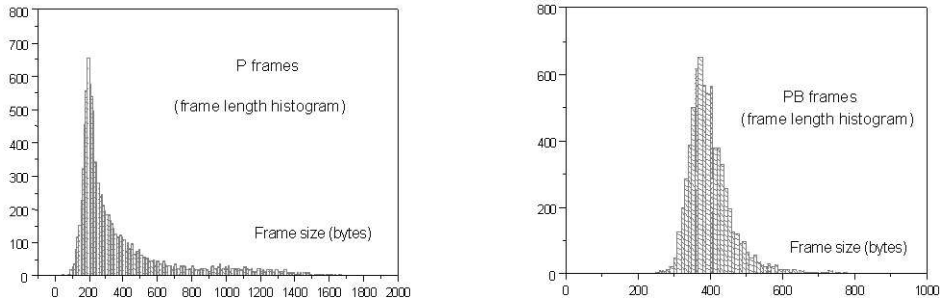

Fig. 3. Histograms of P- and PB-frames for H.263 16Kb/s encoded movie "Mr. Bean".

In Fig. 2 the MF spectrum of the frame size video traces for H.263 en-

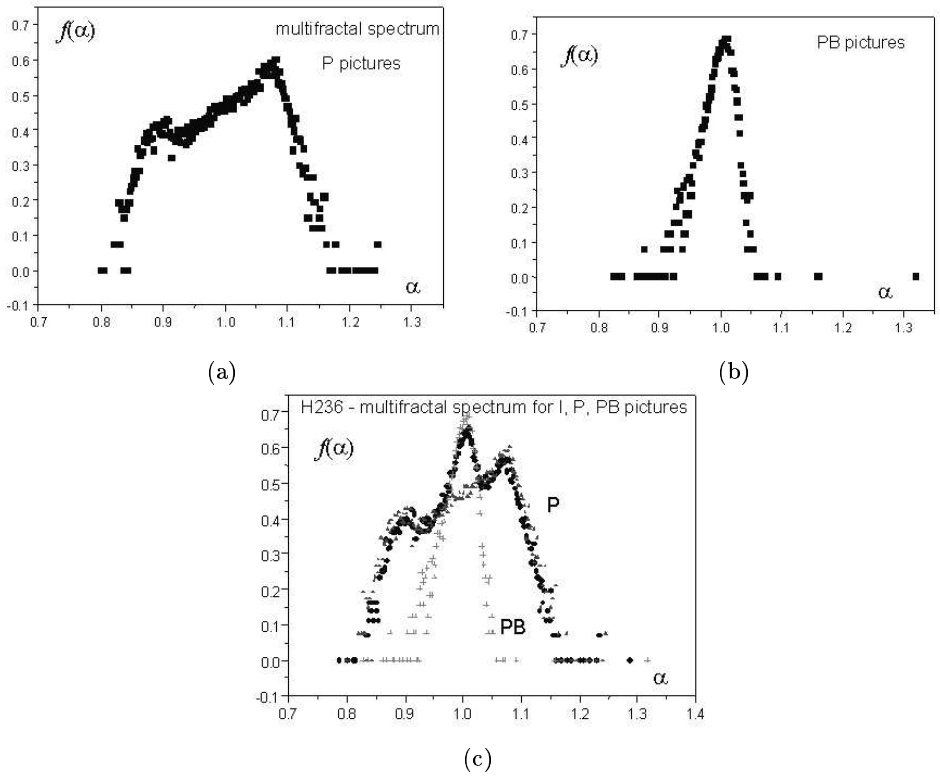

Fig. 4. MF spectra of individual frames for H.263 16Kb/s encoded movie "Mr. Bean", (a) and (b), and their overlapped spectrum (c).

coded movie "Mr. Bean", targeted to bit rate of 16 Kb/s (high compression ratio), are depicted. The compressed video is composed of 17865 frames in total; 7035 PB-frames, 10826 P-frames and only 4 I-frames. Two local maxima, at values of the coarse Hölder exponent α of $\alpha = 1.0$ and $\alpha = 1.1$ are quite observable, as well as less indicative local maximum near $\alpha = 0.9$, indicating to the strong additive process. By detailed analysis we infer that differences in P- and PB-frame sizes provoke this effect.

In Fig. 3 the histograms of P- and PB-frame sizes are depicted. The histogram of P-frames, Fig. 3(a), exhibits global maximum at frame size of 200 bytes and a small local maximum at around 1000 bytes. Conversely, PB-frames have only one significant maximum at 400 bytes, Fig. 3(b). If only P-frames are considered, the MF spectrum as in Fig. 4(a) is obtained, while the PB-frames have the MF spectrum as in Fig. 4(b). The two peaks in MF spectrum for P-frames, Fig. 4(a), correspond to the two different groups of P-frame sizes, Fig. 3(a). By combining P- and PB- spectra, as

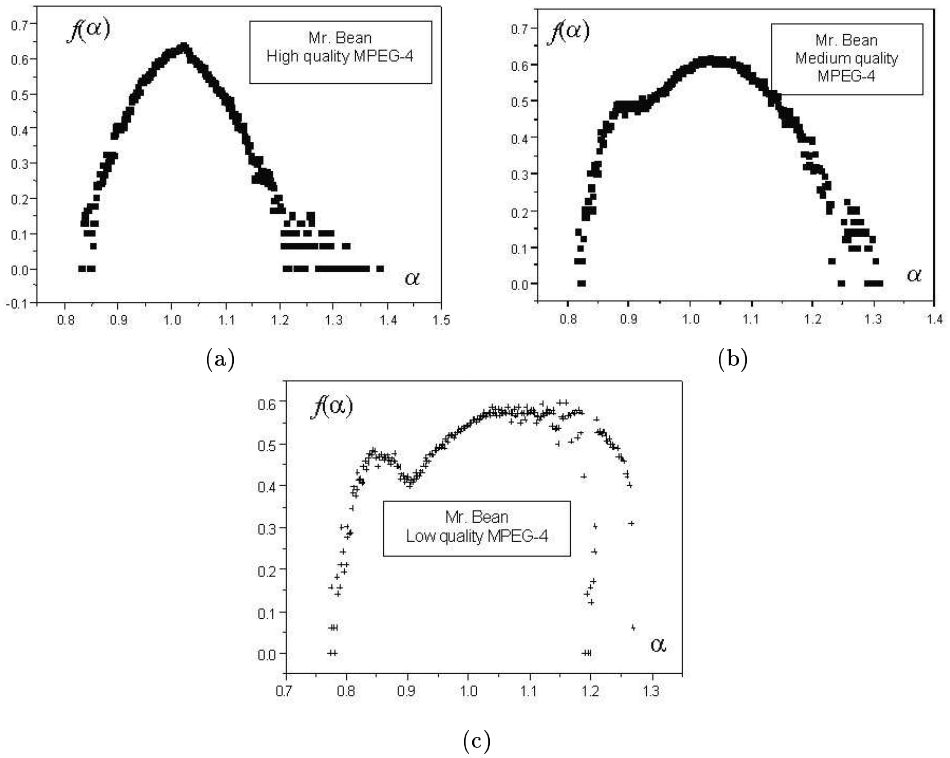

Fig. 5. MF spectra of frame size video traces for MPEG-4 encoded video "Mr. Bean" for different compression ratios: (a) 13:1; (b) 41:1; and (c) 67:1.

shown in Fig. 4(c), the shape as in Fig. 2 is obtained. Note that in this case I-frames have negligible influence to the MF spectrum, since only four I-frames exist in the whole high-compressed video.

The similar results are obtained for MPEG-4 encoded movie "Mr. Bean". In Fig. 5 the MF spectra of the frame size video traces for high-, medium-, and low-quality encoded movie are depicted. As we can see, for high-quality encoded movie (the compression ratio of 13:1), Fig. 5(a), one maximum exists at α near 1.0; for the medium-quality compression (41:1), Fig. 5(b), a second local maximum arises at α near 0.9, while at high compression ratio of 67:1 (low-quality video), Fig. 5(c), the local maximum is more accentuated indicating to the additive process generated by high compression ration and more variability of frame sizes.

More detailed analysis is performed for the video "Mr. Bean." The MPEG-4 compressed low-quality video is composed of 89998 frames in total; 7500 I-frames, 22500 P-frames, and 59998 B-frames. In Fig. 6(a)-(c) the MF

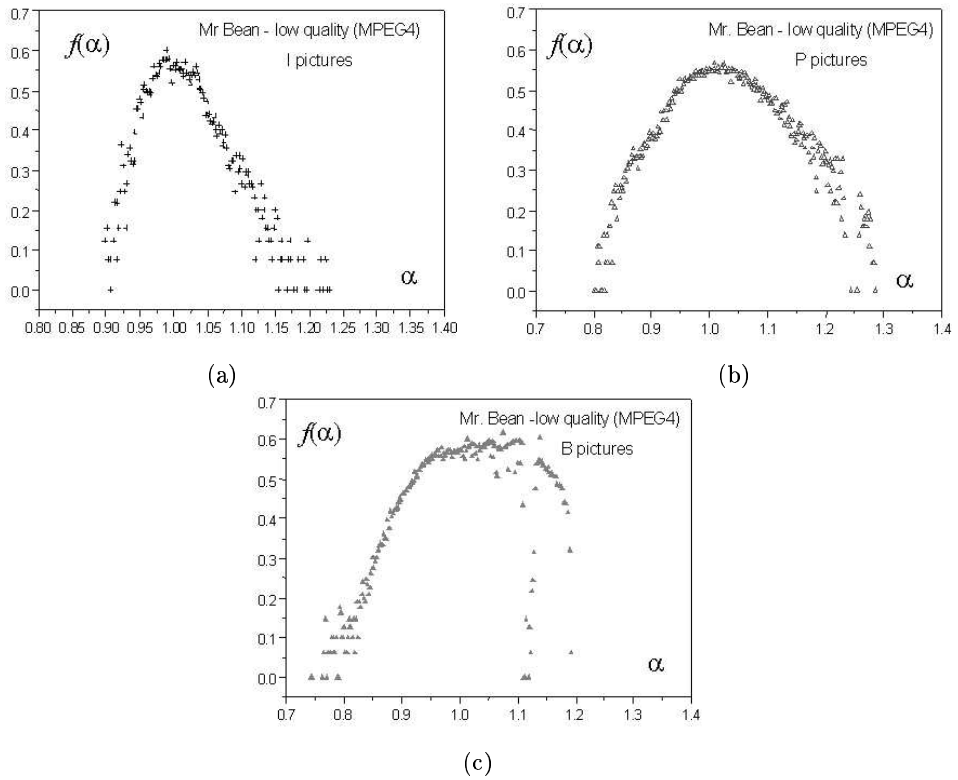

Fig. 6. The MF spectra of I-, P- and B-frame size video traces for MPEG-4 low-quality (compression 67:1) encoded movie "Mr. Bean".

spectra of I-, P-, and B- frame size video traces are depicted. By combining these spectra the shape of MF spectrum for the whole movie, as in Fig. 5(c), is obtained. Certainly, the influence of B-frames on the whole spectrum is predominant, since more than 65% of all frames are just B-frames.

4 Conclusion

The paper considers the fractal and multifractal behavior of compressed video. Video traces in H.263 and MPEG-4 formats, generated at the Technical University Berlin and publicly available, were investigated. It was shown that all compressed videos exhibit fractal (long-range dependency) nature since the Hurst index obtained was in the range between 0.5 and 1.0. Also, it was shown that higher compression ratios provoke more variability of the encoded video stream. This conclusion is approved not only from the

frame size traces [6] but also from the MF spectra of frame sizes. Moreover MF spectra become bimodal as the compression rate increases, indicating to the additive nature of the processes. By analyzing individual frames and their MF spectra the additive nature is approved.

References

[1] J. Russ, *Image Processing Handbook, Third ed.* CRC Press, 2000.

[2] B. Reljin and I. Reljin, "Telemedicine in multimedia environment," in *Telemedicine* (P. Spasić, I. Milosavljević, and M. Jančić-Zguricas, eds.), pp. 22–107, Belgrade: Academy of Medical Sciences of Serbian Medical Association, 2000.

[3] K. Castleman, *Digital Image Processing.* NJ: Prentice Hall, 1996.

[4] A. Netravali and B. Haskell, *Digital Pictures: Representations, Compression, and Standards (Second. Ed.).* Plenum Press, 1995.

[5] K. Rao, Z. Bojković, and D. Milovanović, *Multimedia Communication Systems: Techniques, Standards, and Networks.* NJ: Prentice Hall, 2002.

[6] F. Fitzek and M. Reisslein, "MPEG-4 and H.263 video traces for network performance evaluation," TKN Technical Report TKN-00-06, Technical University, Berlin, 2000.

[7] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson, "Selfsimilarity through high-variability: Statistical analysis of ethernet lan traffic at the source level," in *Proc. ACM*, (Sigcomm), 1995.

[8] M. Taqqu, V. Teverovsky, and W. Willinger, "Estimators for longrange dependence: An empirical study," *Fractals*, vol. 3, pp. 785–788, 1995.

[9] M. Crovella, M. Taqqu, and A. Bestavros, "Heavy-tailed probability distributions in the world wide web," in *A Practical Guide to Heavy Tails: Statistical Techniques for Analyzing Heavy Tailed Distributions* (R. Adler, R. Feldman, and M. Taqqu, eds.), Boston (MA): Birkhauser, 1996.

[10] P. Mannersalo and I. Norros, "Multifractal analysis: A potential tool for teletraffic characterization," (COST257TD(97)32), pp. 1–17, 1997.

[11] I. Reljin, "Neural network based cell scheduling in atm node," *IEEE Communications Letters*, vol. 2, pp. 78–81, March 1998.

[12] I. Reljin and B. Reljin, "Neurocomputing in teletraffic: Multifractal spectrum approximation (invited paper)," in *Proc. 5th Seminar NEUREL-2000, IEEE*, (Belgrade), pp. 24–31, Serpt. 25-27, 2000.

[13] B. Reljin and I. Reljin, "Multimedia: The impact on the teletraffic," in *Book 2* (N. Mastorakis, ed.), pp. 366–373, Clearance Center, Danvers, MA: World Scientific and Engineering Society Press, 2000.

[14] M. Garrett, *Contributions Toward Real-Time Services on Packet Switched Networks.* PhD thesis, Columbia University, NY, 1993.

[15] H.Peitgen, H.Jurgens, and P.Andrews, *Chaos and Fractals.* Berlin: Springer, 1992.

[16] M. Turner, J. Blackledge, and P. Andrews, *Fractal Geometry in Digital Imaging.* Academic Press, 1998.

[17] I. Reljin, B. Reljin, I. Rakočević, and N. Mastorakis, "Image content described by fractal parameters," in *Recent Advances in Signal Processing and Communications* (N. Mastorakis, ed.), pp. 31–34, Danvers, MA: World Scientific Press, 1999.