A Content-Based Optimization of Data Stream Television Multiplex

Jaroslav Polec, Martin Šimek, Michal Martinovič, Elena Šikudová

Abstract—The television multiplex has reserved capacity and therefore we can use only limited number of videos for propagation of it. Appropriate composition of the multiplex has a major impact on how many videos is spread by multiplex. Therefore in this paper is designed a simple algorithm to optimize capacity utilization multiplex. Significant impact on the number of programs in the multiplex has also the fact from which programs is composed. Content of multiplex can be movies, news, sport, animated stories, documentaries, etc. These types have their own specific characteristics that affect their resulting data stream. In this paper is also done an impact analysis of the composition of the multiplex to use its capacity by video content.

Keywords—Multiplex, content, group of pictures, frame, capacity.

I. INTRODUCTION

DYNAMICALLY allocating transmission bandwidth between multiple video streams is one of the major challenges of broadcast systems such as digital video broadcasting (DVB) statistical multiplexing [8] based on joint control of multiple video coders offers a good-quality smoothing solution in this situation. Additionally, an attempt is made to maximize the number of video streams sharing the same transmission channel while facing multiple simultaneous constraints. These are related to bandwidth usage, minimum required video quality, smooth quality transitions over time, and balanced quality between streams [9]. The rate and distortion (R-D) characteristics of the transmitted streams are known as useful tools for solving the resulting constrained optimization problem. Efficient R-D characteristics represent the real behavior of the encoder and so provide a robust control [10].

In [6] the encoder rates are continually adapted. A rate controller determines the channel rate for each encoder as well as the target rate allocated to individual pictures encoded by the encoders. The target rates for individual pictures are computed in order to maintain constant quality across encoders while the encoder channel rates are derived so as to avoid encoder and decoder buffer underflow and overflow.

A lexicographic framework for bit allocation to the multiplexing of multiple VBR video streams onto a single CBR channel is realized in [7]. In the lexicographic framework, the maximum distortion is minimized, then the second highest distortion, and so on, resulting in nearly constant quality. With a suitably constructed multiplexing model, we show that the multiplexing problem reduces to a single-stream CBR bit allocation problem, to which we apply the lexicographic framework. This method has applications for video servers, especially for near-video-on-demand.

In [11] is presented a bandwidth allocation scheme based on content complexity to equalize the overall video quality, in effect a form of statistical multiplexing. Complexity metrics serve to estimate the appropriate bandwidth share for each stream, prior to distribution over a wireless access network. These metrics are derived after entropy decoding of the input compressed bit-streams, without the delay resulting from a full decode. Fuzzy logic control serves to adjust the balance between spatial and temporal complexity metrics. The paper examines constant and varying bandwidth scenarios. Experimental results show a significant overall gain in video quality in comparison to a fixed bandwidth allocation.

In [12] is used Wavelet transform for characterizing measured Ethernet traffic. Research is concentrated on self-similarity presence and show that this phenomenon can be found in measured data.

Those contributions do not deal with position of I, P, B frames in the video. It also does not deal with the mutual position of I frames between videos in multiplex. If we consider the optimization of videos bit rates in multiplex using VBR is strongly dependent on the desired video quality, it is necessary to optimize the mutual position of I frames in a multiplex. Based on the assumption that for the television is in the most cases prescribed uniform length GOP, it is not possible to use any of optimizations that determine the location of frame into the place of sharp cut. I frames are therefore placed at specified times. And these positions have to be optimized in order to avoid either the large fluctuations in the resulting bit rate of multiplexor to fluctuations in the quality of videos of I frames and shortly thereafter.

The optimization of the total flow is also dependent on the movement and activities of spectral images in a multiplex.

A. Extracting Semantics from Video

The semantic information is an important clue in content-based video summarization and subsequent retrieval. With the semantic gap still being unsolved, the research in this area is

J. Polec is with the Institute of Telecommunications at Faculty of electrical engineering and information technology, Slovak University of Technology, Ilkovičova 3, 812 12, Bratislava, Slovak Republic (phone: +421268279409, email: polec@ktl.elf.stuba.sk).

M. Šimek is with the Institute of Telecommunications at Faculty of electrical engineering and information technology, Slovak University of Technology, Ilkovičova 3, 812 12, Bratislava, Slovak Republic.

M. Martinovič is with the Department of Telecommunications at Faculty of electrical engineering and information technology, Slovak University of Technology, Ilkovičova 3, 812 12, Bratislava, Slovak Republic (email: martinovic@ut.fei.stuba.sk).

E. Šikudová is with Department of applied informatics at Faculty of mathematics, physics and informatics at Comenius University in Bratislava.

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:7, No:10, 2013

very vivid. In videos different modalities can help to bridge the semantic gap. These modalities include visual, audio or textual clues, motion. A survey of different fusion technologies of these modalities can be found in [14]. In [15] is created a system that explores complementary modalities of Visual-Color, Visual-Region, Visual-Texture, Audio-Perceptual, Audio-Cepstral and Textual.Low-level features used to extract sematic information include average brightness, dominant color, Local Binary Patterns (LBP), color layout, FFT, shape recognition, color histograms, optical flow, etc. Medium-level features utilized in determination of the semantic information are face detection, motion detection, human detection, salient region detection, speech recognition, speaker identification and gesture detection. Among the best performing detection/recognition methods are SVM, HMM, bag of (visual) words and Bayesian inference.

Results of semantic categories detection are still not satisfactory. In TRECVid 2006 the highest ranking semantic concepts were Weather (~0.65), Sports and Maps (~0.5) [16]. On the other hand the shot boundary detection task in TRECVid 2005 resulted in Precision/Recall of 0,936/0,949 for cuts and 0.788/0.791 for gradual transitions [17].

II. PROPOSED ALGORITHM

Decoder has to get frame I to starting of video play. Frame I is many times bigger than P and B frames and if we sent at the same time frame I, we would unnecessarily waste bandwidth and in worse case we would not have even enough bandwidth. This can be easily avoided by individual video delay of several frames and in order to prevent of frames I overlapping in multiplex.

Algorithm in Fig. 1 works on a very simple principle: At the beginning of algorithm we take the first video without shift. Also we take another video with initial shift set to 0 (i.e. without shift). The following procedure is counting of bit rates in each time and finding maximum of them and this maximum is saved. Next step is increasing the shift by 1 and repeating the previous process (finding a maximum and then comparing with previous one). If current maximum is smaller than previous one, we will record only current maximum as reference value for the next steps and also in which shift is achieved this maximum. Shift is increased again, until the set value of maximum shift is achieved. In our case it is 30. We chose this threshold because the GOP size for DVB-T with H.264 encoder is also a size 30 and shifts are searched like this for its entire length. If GOP takes just over one second it run 25 frames per second and contains 30 frames. If we increase this value, we will achieve a lower bit rate, but at the expense of delay (it could be a very unpleasant especially for sport events: our neighbor watch non delayed video, but our multiplex is delayed for example about three second, so our neighbor would have shouted goal and we would still watch only beginning of goal situation). After achieving the last shift is remembered last maximum and shift in which is occurred this maximum and the next video is taken. This process is repeated until the entire multiplex is filled by the number of streams.
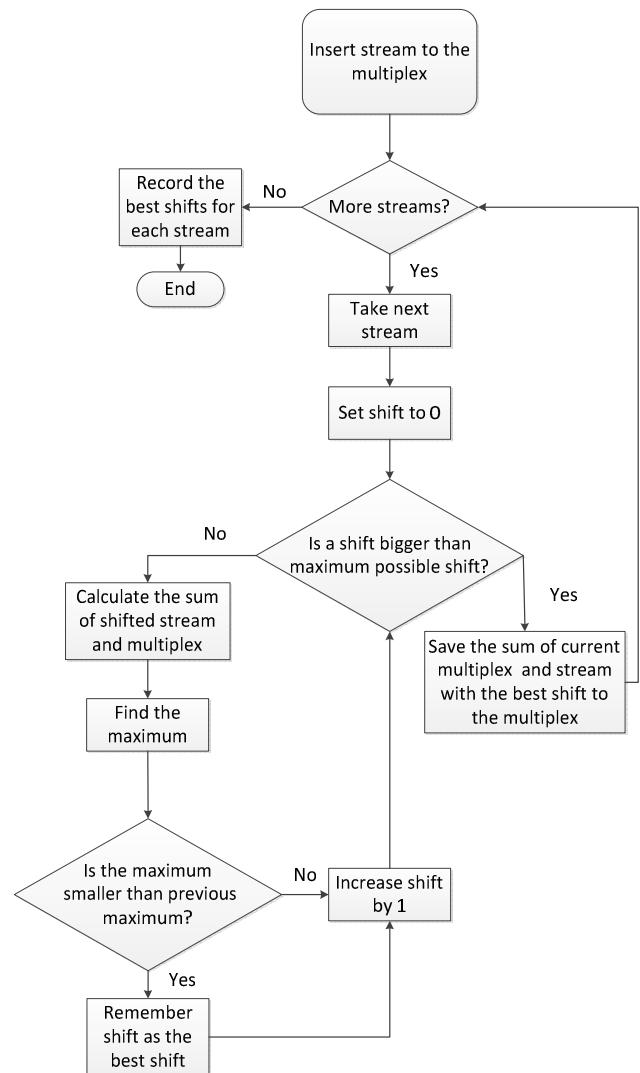

Fig. 1 The flowchart of the algorithm for finding the optimal shift of frames to achieving the lowest bit rate in the multiplex

III. EXPERIMENT

The experiment is realized on a series of videos from Technical University of Munich [1], which are freely available via ftp server for non-commercial purposes. The database contains movies in standard and high definition. Videos are downloadable in the form images (TIFF format). It allows lossless storage. Specifically, these following video sequences:

- carving3 – approximating shot of shop with two people inside
- cc1-cc4 – file contains a video with mass start of cross country skiers captured by lateral camera, process of cross-country skiing captured by back camera and with zoomed camera, camera is static in all three cases, camera of the last video captures skiers from moving car
- forest1 – forest in winter, captured from a moving car, the rising sun from trees
- soccer1 – ball on the field, camera moves gradually up with move of player, in the beginning is the detailed

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:7, No:10, 2013

shotand gradually to the end at length
- soccer2 – the camera is followed the ball, which is kicking by goalkeeper
- soccer3 – camera is placed behind football goal and it shoots the attack
- soccer4 – tilted camera follows the training of football players
- tum1-3 – slowly approaching and isolate shot on the Technical University in Munich and its surrounding captured by above camera
- okt1 – Shots from Oktoberfest which capture carousels at night, lots of lights and colors
- okt3 – shot on balloons and crowd of people
- okt4 – passing people are captured by static camera
- okt5 – shot of carousel over the day

All of these video sequences are in full HD (1080 lines and color subsampled by 3:1:1 system). We converted them by x.264, using color subsampling 4:2:0 and 4:2:2. We recorded output values of PSNR for each components Y, U, V from both types and the resulting bit stream of each video. Brightness values Y were with these ways almost the same. It results from way of subsampling, because only color components are reduced (not brightness). So we display only PSNR for U, V components and bit complexity. All of these values are shown in Table I.

TABLE I
COMPARISON OF PSNR BIT COLOR COMPONENTS AND BIT COMPLEXITY FOR COLOR SUBSAMPLING OF TYPES 4:2:0 AND 4:2:2

	4:2:0			4:2:2		
	PSNR [dB]		Bitrate	PSNR [dB]		Bitrate
	U1	V1	[Mbps]	U2	V2	[Mbps]
cc1	45,875	45,791	22,261	43,485	43,801	25,175
cc2	44,343	45,359	23,315	42,701	44,087	26,359
cc3	43,399	45,423	27,412	42,064	44,407	30,617
cc4	47,508	48,018	24,390	44,015	44,627	26,043
forest	45,595	46,173	34,995	42,893	44,305	37,928
soccer1	44,644	47,435	26,741	43,174	46,116	29,416
soccer2	46,680	48,602	13,512	44,052	46,517	15,400
soccer3	47,127	48,140	23,673	44,751	46,365	25,326
soccer4	46,557	49,035	18,523	44,593	47,862	20,013
okt1	45,076	47,163	16,172	45,411	47,541	18,323
okt3	48,257	48,441	13,270	48,871	48,919	14,268
okt5	45,481	46,417	16,875	46,141	47,203	18,858

As we can see in Table I, the difference of values for PSNR of the color components is in the range from -0,786 dB to 3,493 dB and the average percentage increase in the bit stream is only from 6% to 14%. Due to this fact in the next part of paper, we decided to create multiplexes only of video sequences that have better quality, because they are subsampling by system 4:2:2. Besides these videos we used further videos:
- Aspen – shot of leaves and trees in autumn, slowly move of camera[2]
- ControlledBurn – captures firefighters who sprayed water on the tree and subsequent there is controlled fire of house (burned to the ground) [2]
- RedKayak – captures waterman on the whitewater from near and far away distance [2]
- RushFieldCuts – captures the crowd of people on American football, who after the game rush onto the field [2]
- TouchdownPass – captures passing and subsequent touchdown in American football game [2]
- SnowMnt – alpine peaks covered with snow, slow move of camera [2]
- WestWindEasy – shot of grass in the wind, in the left third of image, subtitles are running on a background [2]
- And as two last video sequences we used non-commercial, freely available short animated films:
- ElephantsDream[3]
- BigBuckBunny [4]

A. Multiplex Capacity

According to [5] we can transfer 38.47 Mbit/s payload data stream in one 8 MHz original analogue TV channel. The question is: how many programs can fit into such a capacity. Therefore, we first randomly selected different video sequences. Then we created multiplex. As we can see in Table II, the maximum value of the bit stream using an optimization algorithm for the 8 programs in a multiplex is about 36.5 Mbps. It is only slightly less than the total capacity of the multiplex. This value is however only a clear video stream without sound or teletext. After adding these components, which is 0.5 Mbps on one program, we get the value 40.5 Mbps and it is inconvenient for our multiplex.

Based on Table II, when we gradually reduce the number of programs in the multiplex, we can reach the maximum number of videos equal to 5 for multiplex without shift and the maximum number of videos is equal to 7 for multiplex with optimum shift.

TABLE II
COMPARISON OF MAXIMUM BIT STREAM AND PERCENTAGE INCREASE FOR MULTIPLEX FORMED FROM 3-8 DIFFERENT PROGRAMS WITHOUT SHIFT AND WITH OPTIMAL SHIFT

	Number of videos in multiplex					
	3	4	5	6	7	8
	maximum [Mbps]					
without shifts	22,9	28,3	34,1	42,1	45,0	62,0
optimal shifts	11,4	18,6	22,9	26,9	28,6	36,5
without shifts/ optimal shifts [%]	200,0	152,4	148,5	156,1	157,1	169,9

B. Multiplex Analysis

In this section we analyze multiplexes by the content of programs in a multiplex. We created programmatically same multiplexes: with slow movement, animated and sports. We created multiple-optimized data flow in the end.

Multiplex of 7 slow, sometimes almost static images, is composed of videos: SnowMnt, ControlledBurn, WestWindEasy, carving3, Aspen a okt4. Videos are described above.

The resulting data stream is shown in Fig. 2 and statistics are given in Table III.
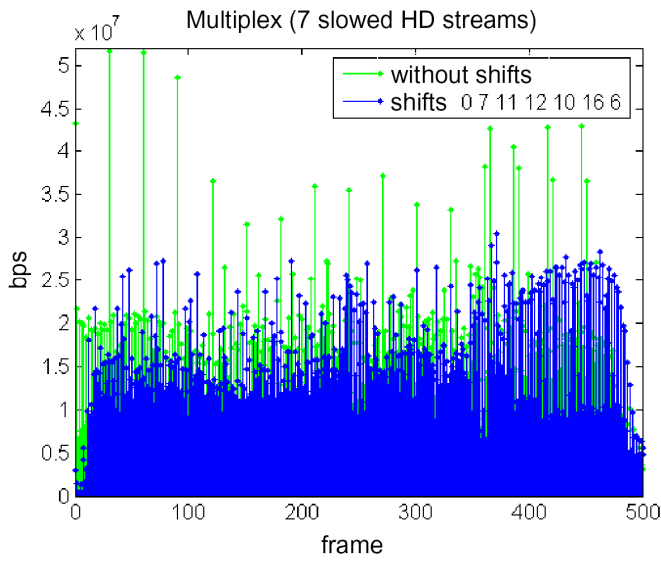
World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:7, No:10, 2013

Fig. 2 Graphical representation of the multiplex for 7 slow programs in high definition arranged without shift and with optimum shift

Fig. 3 Graphical representation of the multiplex for 7 animated programs in high definition arranged without shift and with optimum shift

TABLE III
COMPARISON OF BIT COMPLEXITY AND PERCENTAGE INCREASE FOR MULTIPLEX MADE UP OF 7 PROGRAMS WITHOUT SHIFT AND WITH OPTIMAL SHIFT

	minimum [Mbps]	maximum [Mbps]	mean [Mbps]	deviation [Mbps]
without shifts	2,53295	51,66500	15,38685	7,01766
optimal shifts	4,94065	30,45172	15,29632	3,94693
without shifts/optimal shifts [%]	51,26754	169,66198	100,59186	177,80067

TABLE IV
COMPARISON OF BIT COMPLEXITY AND PERCENTAGE INCREASE FOR MULTIPLEX MADE UP OF 7 ANIMATED PROGRAMS WITHOUT SHIFT AND WITH OPTIMUM SHIFT

	minimum [Mbps]	maximum [Mbps]	mean [Mbps]	deviation [Mbps]
without shifts	2,42817	37,92022	9,07691	4,62019
optimal shifts	1,93615	24,49780	9,20853	4,31482
without shifts/ optimal shifts [%]	125,41254	154,79032	105,57692	107,07717

Because the multiplex is made up of multiple slow scenes or scenes where the only movement is caused by movement of the camera, one would expect that this type will be the least difficult to bit stream. But the opposite is true. The maximum bit rate is 30.45 Mbps compared to 24.49 Mbps for animations (Table IV). Also the average value is 15.29 Mbps to 9.2 Mbps, which is quite a significant difference. It is caused by wide of spectrum of animated videos. Even very slow scenes are more difficult to bandwidth to comparison with very fast animated scenes.

In the next section, we used the before mentioned two animated movies. We cut them to sequences of size 20 seconds. Selected sequences are calmer but also with very demanding movement.

The resulting data stream is shown in Fig. 3 and statistics are given in Table IV.
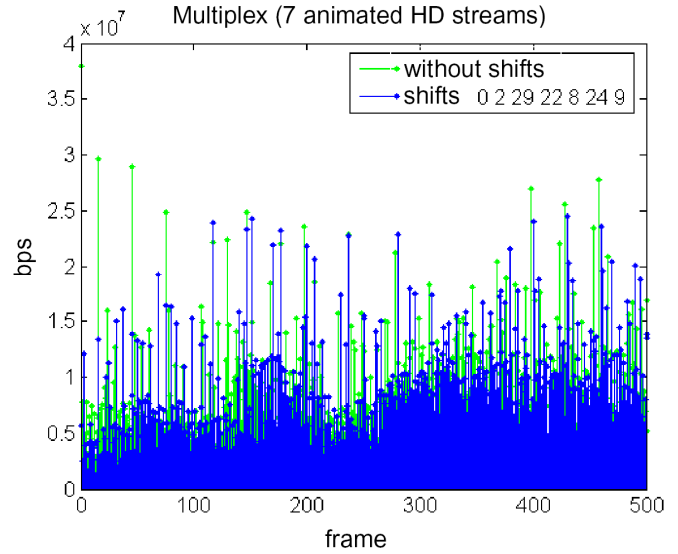
Animated images are significantly different from the real images. They are unique, because their color range is not as wide as in the real shots. It is the reason why the bit complexity is much smaller too. It is the true also in slow scenes, as well as very fast demanding scenes and also the fluctuations in bit rate are lower. It can be seen in Table IV. Maximum bit complexity of multiplex with 7 videos without shift is only 37.9 Mbps. It is significantly less than 52.25 Mbps for sports multiplex. In the case of optimal shift, the maximum bit complexity is only 24.49 Mbps for animated programs. It is the lowest value of all studied types of videos. Composition of multiplex with 7 animated programs therefore very inefficient uses available bandwidth.

Multiplex for 7 sports programs consists of the videos: TouchdownPass, RedKayak, soccer2, RushFieldCuts, cc4, soccer3 and cc1 (each one is described above). These are records of American football match, from whitewater, football and the race of cross country skiers. The resulting data stream is shown in Fig. 4 and statistics are given in Table V.

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
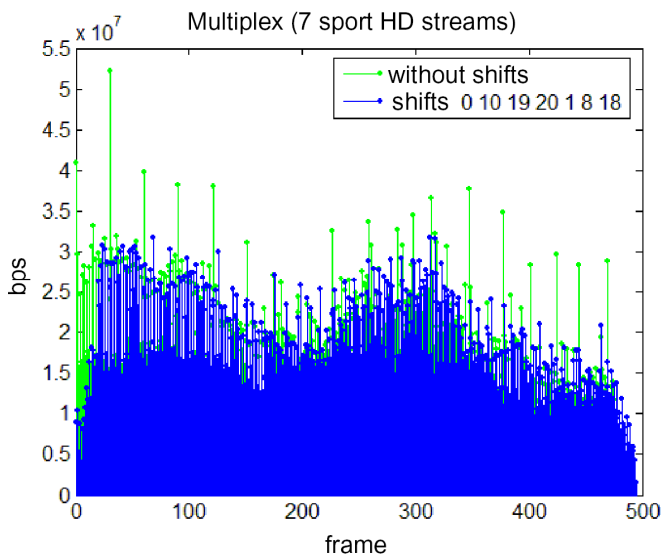Vol:7, No:10, 2013

Fig. 4 Graphical representation of the multiplex for 7 sports programs in high definition arranged without shift and with optimum shift

TABLE V
COMPARISON OF BIT COMPLEXITY AND PERCENTAGE INCREASE FOR MULTIPLEX MADE UP OF 7 SPORTS PROGRAMS WITHOUT SHIFT AND WITH OPTIMUM SHIFT

	minimum [Mbps]	maximum [Mbps]	mean [Mbps]	deviation [Mbps]
without shifts	10,94290	52,25242	20,85599	5,94037
optimal shifts	12,15521	37,75561	20,64093	4,90298
without shifts/ optimal shifts [%]	90,02649	164,54546	101,04192	121,15825

Sports programs are characterized by particularly fast movement and dynamic nature of changes in the image. Of course it does not apply for all kinds of sports, such as pool, chess or darts, but mostly it is true. Similarly, it also applies to all sports videos, we worked with. Therefore, the maximum bit rate is greater than for other programs. In this case value for optimal shift is 37.75 Mbps, which exceed out capacity. If we formed multiplex of 7 sports videos, it would often be at the expense of quality.

By a suitable composition of multiplex by content can be achieved maximum efficiency while maintaining bandwidth utilization claims to the video quality in all programs of multiplex.

Multiplex of 7 different videos consists of videos: tum1, cc1, okt1, aspen, forest1, RushFieldCuts a BigBuckBunny. Every mentioned video is described above.

An important think is, that it contains videos with different complexity on the data flow:
- With minimal movement (tum1 and aspen)
- With a lot of small movement (okt1 and RushFieldCuts)
- With a big changes in movement (cc1 and forest)
- Animated movie in form of BigBuckBunny.

The resulting data stream is shown in Fig. 5 and statistics are given in Table VI.
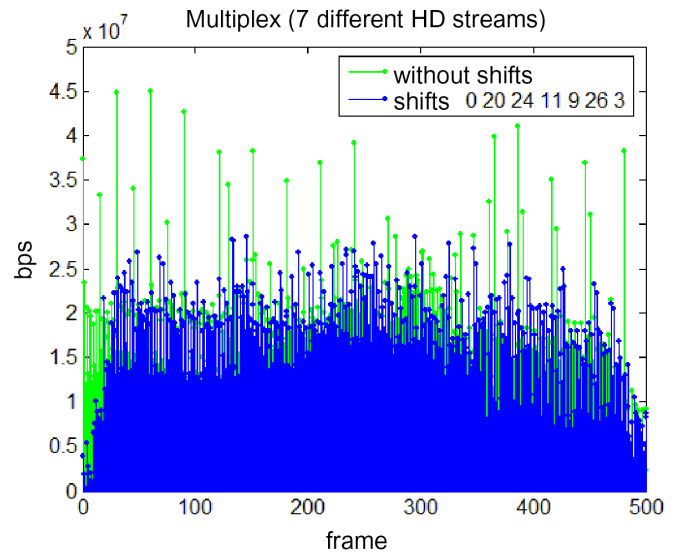

Fig. 5 Graphical representation of the multiplex for 7 different programs in high definition arranged without shift and with optimum shift

TABLE VI
COMPARISON OF BIT COMPLEXITY AND PERCENTAGE INCREASE FOR MULTIPLEX MADE UP OF 7 DIFFERENT PROGRAMS WITHOUT SHIFT AND WITH OPTIMUM SHIFT

	minimum [Mbps]	maximum [Mbps]	mean [Mbps]	deviation [Mbps]
without shifts	10,32945	45,01292	18,84827	5,71381
optimal shifts	10,37830	28,64905	18,54438	3,90069
without shifts/ optimal shifts [%]	99,529306	157,11838	101,63872	146,48185

As we can see, we get the maximum bit rate 45 Mbps without shift with the number 7 videos and it greatly exceeds our capacity. But when we use the optimal shift of frames, we can get only 28.6 Mbps, which is saving. After adding remaining necessary items to multiplex we get value 32.14 Mbps.

IV. CONCLUSION

In this paper we presented a simple algorithm to optimize capacity utilization multiplex. We also made analysis about impact of the multiplex composition according to video content to use its capacity.

Firstly, we created multiplex with a variable number of channels (from 3 to 8). From selected parameters of multiplex: bandwidth 8 MHz and 64 state quadrature modulation, we found that to multiplex we fit 7 videos with optimum shift and only 5 videos without shift. Using optimization algorithm, we saved a significant amount of multiplex capacity and we managed to still get into it by 2 – 3 videos more in full HD.

In next section we passed multiplexes by content: various sports, animated and slow video sequences. All of them consisted of 7 videos. In terms of the multiplex maximum bit rate sports events reached the highest value because only the video stream reached 31.75 Mbps. Second toughest was video completed of slow video sequences with 30.45 Mbps, followed by a variety of content with 28.65 Mbps and

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:7, No:10, 2013

animated video sequences were last with only less than 24.45 Mbps. It should be noticed that all multiplexes are fitted in the desired capacity. A better view of the complexity with different types gave us average bit rate values. Sports multiplex is the most challenging again with 20.64 Mbps, but change is in other rankings: the second in order with 20.64 Mbps are video sequences, followed by slow video sequences with 15.29 Mbps and a significant difference is the smallest average bit rate for animated movies with average value only 9.2 Mbps. Sports videos were a very dynamic with great movement. It showed that they are clearly the most demanding from the given videos. Followed by videos of different content (each of video types contained it). But the most essential difference is between animated and slow movies, where it would seem that slow videos are not difficult, as there is minimum movement. Neither animations is very complex on data stream, because they contain a much smaller range of colors than any other and in addition they have a much smoother edges than real filmed videos.

We could maximize efficiency of our algorithm by supplementing mentioned algorithm for optimizing data flow by moving the GOP in the beginning of each video about automated sorting videos by content [13].

ACKNOWLEDGMENT

REFERENCES

[1] TechnischeUniversitätMünchen, Institute for Data Processing, *TUM Multi Format Test Set*. [Online]. Available: www.ldv.ei.tum.de/videolab, 2011.
[2] Video sequences are owned by NTIA/ITS, an agency of the U.S. Federal Government, Project Number 3141012-300, Video Quality Research, 2008.
[3] Blender Foundation / Netherlands Media Art Institute Available: www.elephantsdream.org, 2006.
[4] Blender Foundation / www.bigbuckbunny.org, 2008.
[5] EN 300 429 V1.2.1,*Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for cable systems*, European Broadcasting Union, ETSI, Sophia AntipolisCedex, April 1998.
[6] Balakrishnan, M., Cohen, R., "Global Optimization of Multiplexed Video Encoders", *Proceeding ICIP '97 Proceedings of the 1997 International Conference on Image Processing (ICIP '97)*. Vol. 1., Santa Barbara 1997, pp. 377 - 380.
[7] Hoang, D.T., Vitter, J.S., "Multiplexing Vbr Video Sequences Onto A CBR Channel With Lexicographic Optimization", *Proceeding ICIP '97 Proceedings of the 1997 International Conference on Image Processing (ICIP '97)*. Vol. 1., Santa Barbara 1997, pp. 369 - 372.
[8] He, Z., Wu, D.O., ''Linear Rate Control and Optimum Statistical Multiplexing for H.264 Video Broadcast", *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1237–1249, 2008.
[9] Changuel, N., Sayadi, B., Kieffer, M., ''Predictive Control for Efficient Statistical Multiplexing of Digital Video Programs'', in *Proc. IEEE Int. Packet Video Workshop*, 2009, pp. 1–9.
[10] Changuel, N., Sayadi, B., Kieffer, M.,"Statistical Multiplexing of Video Programs",*IEEE Vehicular Technolology Magazine*, September 2009, pp. 62 - 68.
[11] Moiron, S., Razavi, R., Fleury, M.,Ghanbari, M.,"Statistical Multiplexing of Transcoded IPTV Streams Based on Content Complexity", *Mobile Multimedia Communications, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, Vol. 77, Springer 2012, pp 60-73.
[12] Kocisky, M., Vargic, R., Kotuliak, I., "Wavelet-domain based estimation of self-similarity in signals", in *Proceeding of RTT 2006*, pp. 441-444.
[13] Šikudová, E.,*Possibilities of Automatic Image Data Classification*, PhD. Thesis, UK Bratislava, 2006.
[14] Atrey, Pradeep, K., Anwar, M.:"Multimodal fusion for multimedia analysis", *Multimedia systems*,vol. 16, Springer-Verlag 2010, pp. 345-379.
[15] Gulen, Elvan, Turgay, Y., Yazici. A. "Multimodal Information Fusion for Semantic Video Analysis"*, International Journal of Multimedia Data Engineering and Management* (IJMDEM) 3.4, 2012, pp. 52-74.
[16] Smeaton, A. F., Wessel Kraaij, P. O." High-Level Feature Detection from Video inTRECVid: a 5-Year Retrospective of Achievements", *Published in A. Divakaran (ed.), Multimedia Content Analysis, Signals and CommunicationTechnology*, DOI 10.1007/978-0-387-76569-3 6, SpringerScience+BusinessMedia, LLC 2009, pp. 151-174.
[17] Smeaton, A. F., Over, P., Doherty, A.R.,"Video shot boundary detection: Seven years of TRECVid activity",Computer Vision and Image Understanding, vol. 114, 2010, pp. 411-418.