Research for Development of Value Added Services for Connected TV

Tanushyam Chattopadhyay
Innovation Lab, Kolkata
Tata Consultancy Services
India
t.chattopadhyay@tcs.com

ABSTRACT

Recent trends on emerging market [1], [2] shows that the connected TV is becoming very popular in developing countries like India, Philippines. Connected TV can be described as an Internet enabled TV. One such product, referred as Home Infotainment Platform (HIP) [3], combines the functions of a television and a computer, by allowing customers to use their television sets for low-bandwidth video chats and access internet websites. It is now commercially available in India [4] and Philippines [5]. The research presented in this thesis is motivated by the business need to implement some value added services, hereinafter referred as *VAS*, on top of this product which can work as the Unique Selling Point for the above mentioned product. Main motivation behind the research described in this thesis is to develop some interesting VAS for the product of the company using frugal computing. As a result of that the thesis is not very focused to a particular problem, instead provides a frugal solution for several goals need to be achieved to develop the VAS for the global product. Some of such planned VAS for those connected TVs are video conferencing, Video encryption and watermarking, context based web and TV mash up, video summarization and Electronic Program Guide (EPG) for cable feed channels [8]. This thesis is devoted to the development of those above mentioned VAS for connected TV. As all these services need to be developed on an embedded platform, the primary task is to realize the required video CODEC on the target DSP platform. As H.264 is adjudged as the best video CODEC of the day [9] because of its compression efficiency, video quality, and network friendliness, we have developed most of the VAS on top of H.264 CODEC. Security can be ensured by either encrypting the video or putting a watermark in the video. Context can be extracted at top level by recognizing the channel the user is viewing and then getting the relevant information from the website of that particular channel. On the other hand, textual information in a TV show provide some information related to the show at any particular instant of time.

1. INTRODUCTION

The total media viewing and sharing experience is changing and getting richer everyday, as videos, music and other multimedia content flood the Internet. The main reason behind the popularity of this product (HIP) in developing countries is that the Internet penetration in those countries is significantly low [6] compared to the penetration of Television. As per the report [7], 75% of the population of India owns a TV. TV has been a favored device of home infotainment for decades. In order to provide an unified Internet Experience on TV, it is imperative that the Internet experience blends into the TV experience. This in turn means that it is necessary to create novel VAS that enrich the standard broadcast TV watching experience through Internet capability. This necessity is eventually translated into the need for different applications like secured distribution of multimedia content, communication using video chat over TV, and applications that can understand what the user is watching on broadcast TV (referred to as TV-context) and provide user with additional information/interactivity on the same context using the Internet Connectivity. Understanding the basic TV context is quite simple for digital TV broadcast (cable or satellite) using metadata provided in the digital TV stream. But in developing countries, digital TV penetration is quite low. For example, in India, more than 90% TV households still have analog broadcast cable TV. Understanding the TV context in the analog broadcast scenario is really a big challenge. Even for the small percentage of homes where satellite TV has penetrated in form of Direct-to-Home (DTH) service, almost all of them lack in back-channel connectivity for proving true interactivity.

2. REVIEW OF RELATED WORKS

In this section a brief overview of the state of the art in the related field is given. *Realization of Video CODEC on Embedded Platform :* In the literature [9] it is reported that the improvement in video quality and compression ratio for H.264 is obtained at the cost of increase in computational complexity and memory requirement. So the State of the art is analyzed in light of these challenges. Detail discussion can not provided here because of the page constraint and so we are providing the gist of the analysis here. *Reduction of Computational Complexity :* Two different approaches were taken to reduce the computational complexity which can

be measured in terms of Mega Cycles Per Second (MCPS). These approaches are (i) Platform independent optimization and (ii) Platform specific optimization. Some approaches are also there who had described the optimization of the encoder execution time as a whole like [10], [11]. *Memory Optimization:* In [12] authors have proposed a novel near-optimal filtering order so that significant reduction in memory requirement is achieved. This work also gives significant reduction reduces MCPS. However, their methodology is applicable to an FPGA prototype. It cannot be used in a commercially available DSP platform, where the user does not have the flexibility to modify the hardware architecture. The above state of the art reveals some limitations like the platform independent optimization techniques gives good optimization but at the cost of coding efficiency. Moreover some of these algorithms are sub optimal and not compliant to the standard. These algorithms are generic and thus can be applied to any type of videos. As the target application is mainly video telephone and video conference, the motion in the videos are very less. So if the nature of video can be exploited, more optimization can be obtained. Thus in this thesis an optimization technique is proposed based on the statistical analysis of the selected mode for these type of videos. On the other hand, the platform specific optimizations are not suited for our choice of video conferencing/video telephone. No literature also focuses on efficient rate control at low cost to get a better video quality.

A comprehensive survey on the H.264 video security is described in [13], [14]. As the video security itself is a vast field of research, we have restricted the State of the Art analysis to the study of encryption and watermarking for videos and more specifically for H.264 compressed domain videos. *Encryption:* Description of video encryption can be found in [15] - [18]. Some of such techniques are (i) Encrypt the motion vector, (ii) Encrypt the entropy coded stream or (iii) Scramble the prediction modes to achieve encryption. But to the best of our knowledge only two such work can be found in [16], [17]. *Watermarking:* Different classifications for watermarking technology is described in [14]. Broadly video watermarking techniques can be classified in two types of approaches namely (i) Pixel Domain where watermark can be directly inserted in the raw video data and (ii) Compressed domain where watermark is integrated during encoding process or implemented by partially decoding the compressed video data. The major problem of implementing the pixel based approaches in the proposed solution is that there is an additional overhead of decoding the compressed video. Moreover the watermarking technique for the proposed system should be compliant to compressed H.264 video format which differs from the previous video codecs in different aspects as described in [9]. We have also described the differences between H.264 and other video codecs in chapter 2.

Text Information Extraction from Video: The input video format for the proposed system is different for different sources of input signal. The input video may come from a Direct To Home (DTH) service or in form of Radio Frequency (RF) cable. In case of DTH, the input video is in H.264 or MPEG or any other compressed digital video format and on the other hand in case of video RF cable, the input is an analog video signal. In the second case initially the video is digitized for

further processing. The Text Information Extraction (TIE) module localizes the candidate text regions from the video. The state of the art shows that the approaches for TIE can be classified broadly in two sets of solution: (i) Using pixel domain information when the input video is in raw format and (ii) Using the compressed domain information when the input video is in compressed format. A comprehensive survey on TIE is described in the paper [19] where all different techniques in the literature between 1994 and 2004 have been discussed. A survey on the recent works in this field can be found from [20].

EPG for RF Enabled TV: Some related work on the channel logo recognition can be found in [21] - [24]. The best performance is observed in x86 platform for the approaches described in [24]. But the approaches taken in [24] involve PCA and ICA which is very much computationally expensive and thus is difficult to be realized in the said DSP platform to get a real time performance. So there is no solution is available in the literature that can recognize the channel logos realtime and can provide the EPG for RF feed TVs.

3.   MOTIVATION FOR THE PRESENT WORK

Review of the previous works on VAS for HIP like systems reveals that most of the studies concentrate on different sub-problems instead of providing a complete end to end solution. The work embodied in this thesis is motivated to fill this gap. The major challenge in developing such a product is the resource constraint namely CPU speed and memory of the target hardware. Some of these algorithms describe a good solution for some of the sub-problems in PC environment. But these solutions cannot be implemented on a fixed point DSP platform. The proposed study is focused on developing the following VAS like (i) security of the broadcast video, (ii) context information extraction from streamed video. Moreover all of those solutions need to be deployed in the target hardware. So we have a plan to incorporate the following VAS as a feature for the HIP.

Low bandwidth video applications: This feature enables the user to do video conferencing with another person having the similar HIP installed in his/her home while watching TV. The basic motivation behind this feature is that the TV screen would be minimized to a lower resolution and the user can use the rest part of the TV screen for video conferencing. This feature comes as the wish list from the customers of urban area of India whose wards are working abroad. Another such solution based on low bandwidth requirement is place shifting solution. This solution enables the user to access the home video content over broadband. But both of these solutions can be implemented when there is an efficient video CODEC, satisfying the requirement of high video quality at a low bandwidth, is realized on a DSP platform. As H.264 is proved to be the best video CODEC of the day, we have implemented H.264 on a low cost DSP platform.

Video Encryption: This feature was motivated by the demand from the TV Channel agencies when the PVR was set in the market. The video encryption algorithm allows the user to record the video content using the key which can be derived from the hardware identification number of the PVR or HIP. As a consequence the user can be tracked if he/she

wants to use the recorded content for some illegal business purpose.

Video Watermarking: The need for video watermarking was motivated by the need of one of the major content provider company. They had a need to insert watermark to the content provided by them in a content delivery network (CDN). The same algorithm can be extended for streaming video applications like video on demand (VoD) services provided by the DTH service providers. They also looked for a watermarking evaluation system that can evaluate any watermarking system.

Mash up of TV context and Internet information: Living in a generation of Google TV, Yahoo Connected TV, it is impossible to sustain in the market of connected TV without providing the mash up feature. But as in India most of the people are using analog cable TV and all of those above mentioned products are based on Digital videos only, there is a need to develop such system to address this variation of input, too. Moreover the quality of the videos obtained from analog feed video is quite poor in comparison to those obtained from the DTH service.

EPG for RF feed TVs: The same gap in technology arising from the source of video content motivates us to develop such a service of EPG for the users using RF feed signal for TV.

We have proposed the solutions that can perform at per of the 80% accuracy and efficiency of the related best PC based solution at a 20% cost in terms of execution time and hardware cost. This concept, commonly known as froogle computing, is mainly targeted for the CE products in developing countries. Current thesis is mainly motivated to provide such froogle solutions that can be deployed on the top of the HIP product already developed by the organization.

4. SCOPE OF FUTURE RESEARCH

The study presented here can be extended in several directions. Some of them are highlighted below:

- *Video Screen Layout Segmentation*: The layout of a video is very complex. We have tried to run different document page layout segmentation techniques on different video frames of news video. But none of these methods can produce a significant result.

- *Frame by frame annotation of video frame using multimodal cues*: The proposed method for mash up of web information and TV context is based on textual content of video only. But a better result perhaps can perhaps be obtained if multimodal cues like audio and image can be used. This can be used for annotating the frames and indexing the video.

- *Cross lingual Information Retrieval*: The textual content from the news video can be further used to retrieve related information from other languages. Script identification and subsequently Cross lingual Information

Retrieval (CLIR) are further research issues involved with this problem.

- *Automatic channel logo region identification*: We have found that the channel logo region identification for the animated channels is a challenge. Automatic localization for these channels (like 9xM) is a possible future extension of the present research.

5. CONTRIBUTION OF THE THESIS

As far the state of the art is concerned, this thesis has several contributions for development of some VAS for a connected Television. Some of the major contributions are briefly discussed below:

- In this thesis some novel approaches to provide better video quality, coding efficiency and reduced MCPS even under the constraint of target hardware has been proposed. Improvement in video quality and coding efficiency under a constant bit rate is achieved by implementing a novel algorithm for adaptively selecting the basic unit for rate controlling. The proposed method also reduces the computational complexity using platform independent and platform specific optimization techniques and yet meets the very low memory constraint of the target processor for a standard H.264 baseline encoder without sacrificing the rate-distortion performance. The platform independent optimizations are useful as this version of the code can be ported to any DSP platform for further platform specific optimizations. Almost 40% MCPS reduction with respect to optimized reference code is achieved at the cost of less than 1% reduction in Peak Signal to Noise Ratio (PSNR).

- This thesis deals with an encryption scheme for H.264 video that can be implemented on a DSP platform. In case of Personal Video Recorder (PVR) enabled STBs and connected TVs any user can easily store any TV program. The proposed technique is capable to protect illegal distribution of video content stored in PVR. This thesis presents a fast yet robust video encryption algorithm that performs real-time encryption of the video in H.264 format on a commercially available DSP platform. This algorithm is applied in a real-time place-shifting solution on DSP platform, too. However, the approach has no negative effect as far as compression ratio and video quality are concerned. Mathematically, it can be shown that the proposed method is more robust than those methods for encrypting H.264 video described in the state of the art analysis.

- With the advent of high-speed machine, a hacker, now a day, finds it less difficult to break any encryption key even though it may require large number of attempts. Therefore, an encryption method alone is not sufficient for copyright protection and ownership authentication of stored and streamed videos. In this thesis digital watermarking techniques has been proposed for this purpose. A fast method of watermarking of streamed H.264 video data is proposed in the thesis to meet the real time criteria of a streaming video. This solution

was deployed in a content delivery network (CDN) environment, too.

- In this thesis a novel TV and web mash-up application is described. This application initially extract the relevant textual information from the TV video coming in either analog or digital format and then mash up the related information from the web to provide a true connected TV experience to the viewers. Unlike digital TV transmission it is not possible to automatically get contextual information of TV programs from any Meta data. The text in a TV channel is extracted by text region identification followed by pre-processing of the text regions and performing Optical Character Recognition (OCR) on the text regions. The applications are presented for x86 based PC platform and ARM based dual-core platform. This type of system is not available in the literature.

- The thesis presents a novel method for recognizing the channel logos from the streamed videos in real time, which has various applications for VAS in the connected TV space. The prototype is developed in X86 platform and then ported on a commercially available DSP with nearly 100% accuracy in real time. In India, where most of the people are still watching TV using Radio Frequency (RF) feed cable, this image processing based approach solution for providing EPG is novel in nature.

6. REFERENCES

[1] A. Wooldridge, "A special report on innovation in emerging markets", *The Economist*, Page(s) 6, 17th April, 2010.

[2] B. Stelter, "A TV-Internet Marriage Awaits Blessings of All Parties", *http://www.nytimes.com/2011/01/10/business/media/10tv.html?_r=1*, 2, 9th January, 2011. Last accessed on 14th January, 2011.

[3] A. Pal, C. Bhaumik, M. Prashant, A. Ghose, "Home Infotainment Platform", *UCMA2010*, Miyazaki, Japan, June 2010.

[4] Economic Times, "Dialog can raise internet penetration", *Economic Times Kolkata*, Section-Business and IT, Page(s)5, Apr 27, 2010.

[5] "SMART launches SurfTV", *http://smart.com.ph/corporate/newsroom/SurfTV.htm*, Last accessed on 13th Jan, 2011.

[6] "World Internet Usage Statistics News and World Population Stats", *http://www.internetworldstats.com/stats.htm.*, Last Accessed on Oct 2010.

[7] "ITU report baffled over RP's high mobile phone, TV penetration standard document styles", *http://technews.com.ph/?p=1627*, Last Accessed on Oct 2010.

[8] T. Chattopadhyay and C. Agnuru, "Generation of Electronic Program Guide for RF fed TV Channels by Recognizing the Channel Logo using Fuzzy Multifactor Analysis", *ISCE'10*, 7-10 June, Germany, 2010.

[9] I. E. G. Richardson, "H.264 and MPEG-4 Video Compression", *ISBN 0-470-84837-5*, 2003.

[10] X. Kim and C. C. Jay Kuo, "A Feature-based Approach to Fast H.264 Intra/Inter Mode Decision", *ISCAS'05*, Page(s)308-311, May 23-26, 2005.

[11] H. Wang and Z. Zhu, "Fast Mode Decision and Reduction of the Reference Frames for H.264 Encoder", *IICCA'05*, Vol. 6, Page(s) 1040-1043 , June 2005.

[12] S. Y. Shih, C. R. Chang, and Y. L. Lin, "A near optimal deblocking filter for H.264 advanced video coding", *Asia and South Pacific Conference on Design Automation, 2006.*, Page(s) 24-27, 2006.

[13] T. Chttopadhyay and A. Pal, "A Survey on Video Security with Focus on H.264: Steganography, cryptography and watermarking techniques", *Proc. of the 2nd National Conference on Recent Trends in Information System (ReTIS 2008)*, Page(s) 63-67, Kolkata, India, 2008.

[14] S. Bhattacharya, T. Chttopadhyay, and A. Pal, "A Survey on Different Video Watermarking Techniques and Comparative Analysis with Reference to H.264/AVC", *ISCE'06*, Page(s) 616-621, Russia, 2006.

[15] Y. Ye, X. Zhengquan, and L. Wei, "A Compressed Video Encryption Approach Based on Spatial Shuffling", *Proc. of 8th International Conference on Signal Processing*, Volume 4, Page(s)16-20, Greece, 2006.

[16] Y. Li, L. Liang, Z. Su, and J. Jiang, "A New Video Encryption Algorithm for H.264", *ICICS'05*, Page(s) 1121-1124, Thailand 2005.

[17] Y. Zou, T. Huang, W.Gao, and L. Huo, "H.264 video encryption scheme adaptive to DRM", *IEEE Transactions on Consumer Electronics*, Vol.52, no.4, Page(s) 1289-1297, Nov. 2006.

[18] Z. Liu and X. L. Sch, "Motion vector encryption in multimedia streaming", *Proc. of 10th International Multimedia Modelling Conference*, Page(s) 64-71, Australia, 2004.

[19] K. Jung, K. I. Kim, and A. K. Jain, "Text Information Extraction in Images and Video: A Survey", *Pattern Recognition*, Volume 37, Issue 5,Page(s) 977-997, May 2004.

[20] T. Chattopadhyay, A. Pal, A. Sinha, "Recognition of Characters from Streaming Videos", *Chapter 2 of Character Recognition*, Published by SCIYO,Page(s) 21-42, 2010, ISBN 978-953-307-105-3.

[21] E. Esen, M. Soysal, T. K. Ates, A. Saracoglu, and A. A. Alatan, "A fast method for animated TV logo detection", *CBMI 2008*, Page(s) .236-241, June 2008.

[22] A. Ekin and E. Braspenning, "Spatial detection of TV channel logos as outliers from the content", *in Proc. VCIP*, SPIE, 2006.

[23] J. Wang, L. Duan, Z. Li, J. Liu, H. Lu, and J. S. Jin, "A robust method for TV logo tracking in video streams", *ICME, 2006*, ICME, 2006.

[24] N. Ozay and B. Sankur, "Automatic TV Logo Detection And Classification In Broadcast Videos", *EUSIPCO 2009*, Page(s) 839-843, Scotland, 2009.