

Saliency Aware Resource Saving in Hand-Held Devices

P. Yadav, N. Ramasubramanian, A. Joshi and P. Tamizharasan

National Institute of Technology, Tiruchirappalli, India

{1993praveen@gmail.com;nrs@nitt.edu;adj.comp@coep.ac.in; tamizh5500@gmail.com}

Abstract: The emergence of new operating systems and applications for mobile phones and tablets has necessitated the need for power optimization. Storage space has become another matter of concern as new operating systems have started supporting video codec and formats originally meant for desktop application without compression and conversion. The work presented here tries to identify the region of interest for video by combining the approach of feature extraction with natural statistics for dynamic analysis of the scene. The portion outside the region of interest in the original video is depreciated in order to increase redundancy for pixel value in a frame. With the increase in redundancy the file size gets reduced during encoding. While playing these videos quality and power consumption is tested. Because of saliency model, the change in quality is not notifiable. The result shows that the reduction in power consumption and size of the file varies for different video codecs in different samples. On an average, the power consumption and the file size of video are reduced up to 7.6% and 46% respectively. Since usability is a very subjective issue which can vary for different video samples also, the range of variation in power consumption and size is high.

Keywords: *Saliency Map, Video Codecs, Usability Testing, Power Consumption.*

1 Introduction

Hand-held devices are rapidly replacing traditional computational devices in day to day life. They are preferred due to their compact size and ease of mobility. For the very same reason, they are constrained in terms of resources like power, storage space computational power etc. [23]. Various hardware optimizations like Dynamic Voltage and Frequency Scaling [5], power management through dynamic approach in disk and Network Interface Card [19][20] have been proposed for power optimization in these devices. Apart from this there were several algorithmic approaches like middleware and compiler optimization is also giving promising results [3] [18]. Video applications are one of the primarily used applications in handheld devices. They also play a major role in the usability of the device [21].

In our previous work, we have shown that there is a strong correlation between power consumption for video application and codec used for encoding [15]. It is believed that due to algorithmic complexities and different loss ratio the power consumed by video having different video codec is different [26][4]. The human nervous system cannot focus on bulk of information at a time. There are several features on which attention is more compared to others [14]. Saliency mapping is a technique to measure visual attention of the scene. It can be applied to a video sample also to check which part of it gets more attention. This technique has two parts: bottom-up saliency checks if a scene captures attention or not whereas top-down approach is based on the goal for which a person is searching [9]. Video encoding techniques are responsible for reduction of size in videos based on redundancies in a frame and between the frames of a video.

The work presented here is based on the saliency mapping of videos to identify region of attention. The redundancy in remaining part is increased by diminishing the features through partial blurring. This results in the size of the file becoming smaller as compared to the originally encoded file without affecting usability. Further, when power consumption for these samples is measured, it is observed that modified video consumes less power than the originally encoded videos. Since saliency mapping is used for finding the region of attention, the depreciation in the quality of the video is not much observed.

Related work in this field has been described in section 2. Section 3 explains the experimental methodology and the basics of saliency mapping used, while section 4 describes the observed results and analysis.

2 Related Work

It is believed that storage and transfer have a major impact on power consumption. So Koen et al. [6] have applied memory management in parallel processing to achieve low power and area for the QSDPCM video codec. Morgan et al. [11] and Chimienti et al. [1] tried to optimize architecture so that video codec can perform better at low frequency. Atsushi et al. [2] have shown that algorithmic technique can also be helpful for reducing power consumption in video decoding by using software direct memory access (DMA). Mohsen et al. [10] have used complexity aware technique for the H.264 video codec for motion compensation using vector prediction and interpolation which is believed to be efficient for reducing power consumption. Krishna et al. have moved one step ahead by making the encoding distributed for saving power and memory [7].

Peter et al. have used saliency mapping and implemented the selective rendering on the GPU [12]. It leads to reduction in computational complexity with minimal reduction in quality. Barendregt and Bekker utilized usability factors for encoding of video in computer games targeted for children to achieve higher user satisfaction level [22]. Wei et al. have focused on zooming and finding the Region of Interest so that quality of service can be improved for mobile devices where bandwidth is limited [24]. By mixing fidelity adaptation with saliency mapping, Shao et al. have achieved significant compression ratio during encoding [17]. Zhicheng et al. have used saliency based bit allocation for better compression with minimal difference in quality [25]. Ndijik et al. have discussed issues related to saliency based video compression and concluded that bit rate can be significantly saved using it [13]. The work presented here extends these approaches by finding tradeoff between usability and saving resources like power and file size. For this purpose, a combination of natural statistics and feature extraction based saliency map is used.

3 Proposed Method

3.1 Overview

The proposed method consists of three parts. As shown in Fig. 1, in the first part two saliency mapping techniques based on natural statistics and feature extraction are combined to achieve more accurate region of interest. This procedure is very critical because features outside the region of interest are intended to be diminished in later stages. So, wrong perception of Region of Interest may decrease the quality of service beyond acceptance level. In the second part, samples are encoded using different video codecs. These video samples in the final part are compared with original encoded samples which do not have any features diminished on the basis of video quality, power consumption and file size.

3.2 Saliency Map Used

The two approaches for calculating accurate region of interest are combined here. The first approach is based on natural statistics for dynamic analysis proposed by Lingyum et al. which is the extension of Bayseian framework using natural statistics as shown in Fig. 2 [8]. The saliency map generated by this procedure is further processed using the second approach using feature extraction and segmentation proposed by Sebastian and Alvaro [16].

In the natural statistical procedure, bottom up saliency is used for exacting features based on content of scene and overall saliency is generated using source scene and the target to be searched. The saliency is calculated using the equation (1).

$$\log s = -\log p(F = f) = \sum_{j=1}^5 \sum_{i=1}^{15} \left| \frac{f_{i,j}}{C_{i,j}} \right|^{\theta_{i,j}} + constant \quad (1)$$

This formula is based on Gaussian filter. Here $-\log p(F=f)$ is a self-information based saliency for a point f . $F=f$ signifies visual feature at point f . In the equation, i is the index of spatial filters and j is an index for temporal scales. C is the scale parameter and Θ is shape parameter. The output after applying this equation on a frame of a video is a gray scale frame on which next method is applied.

In the next method first on-center and off center difference are calculated using six filters. All of the filters are defined based on single integral image as in equation 2.

$$C = \sigma 2^s \quad (2)$$

Here s represents the scale and σ represents surround. Surround is defined using equation 3 and 4.

$$surround(a, b, C) = rectsum(a - C, b - C, x + C, b + C) - \frac{i(a, b)}{(2C + 1)^2} + 1 \quad (3)$$

Here $rectsum(a1, b1, a2, b2)$ is sum of rectangular area defined by point $(a1, b1)$ and $(a2, b2)$ in layer i . In the next step intensity submap is calculated for each pixel using equation 5 and 6.

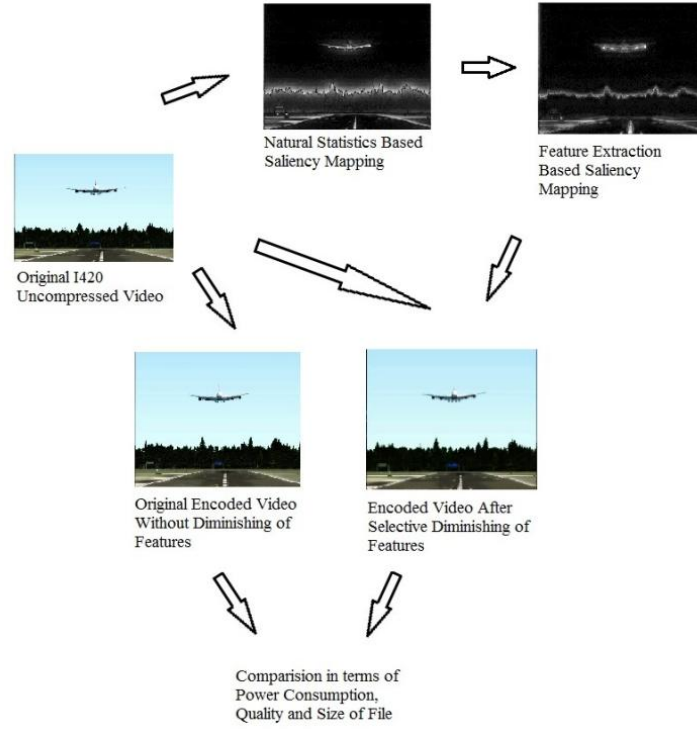


Fig. 1. System Overview for comparison of video samples

$$Int_{on,C}(a, b) = \text{maximum of } \{centre(a, b) - surround(a, b, C), 0\} \quad (5)$$

$$Int_{off,C}(a, b) = \text{maximum of } \{surround(a, b, C) - centre(a, b), 0\} \quad (6)$$

C is the surround, on represents on-center difference and off represents off-center difference respectively. After summing up all submaps final saliency map is developed using equation 7 and 8.

$$Int_{on} = \sum_C Int_{on,C} \quad (7)$$

$$Int_{off} = \sum_C Int_{off,C} \quad (8)$$

4 Experimental Method

4.1 Video Encoding

Different types of videos are used for the purpose of analysis, which includes music videos, part of movies with different themes, videos from computer games. First a video sample is converted into uncompressed I420 encoded format so that exact compression can be compared after encoding. Combination of natural statics and feature extraction algorithm is applied to obtain a salience map on uncompressed sample. The uncompressed video is further encoded into four popular video codecs: : XVID, DIVX, MPEG and DIV3. Before encoding, the area outside the region of interest is blurred to increase redundancies in a frame. This will help in reducing the size of file while encoding.

4.2 Power consumption analysis

The power consumption by different codecs is measured on Aakash Ubiplate 7C+ android tablet using Little Eye performance analysis and monitoring tool. This tool measures power on the basis of a power model which is available in two categories.

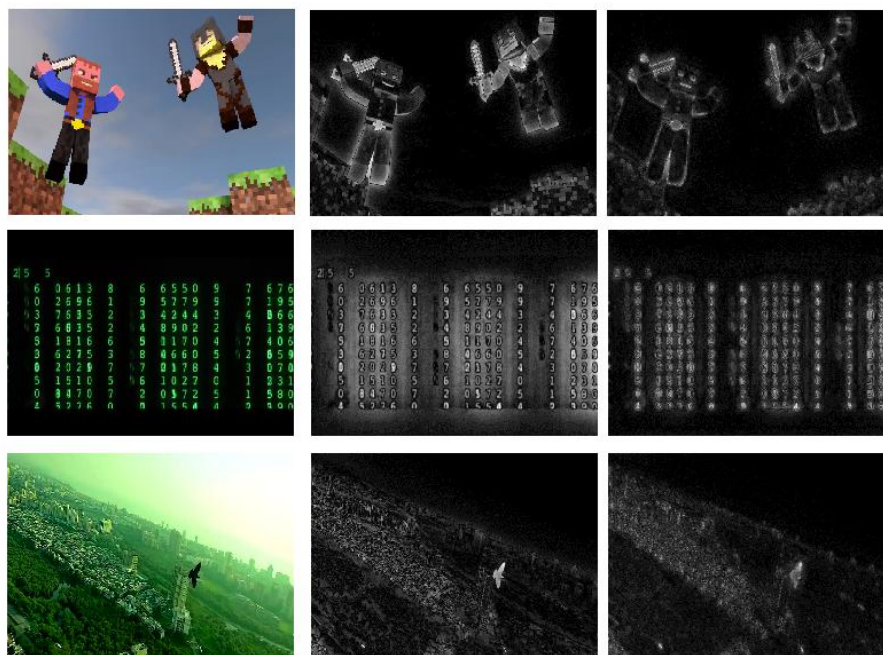


Fig. 2. The saliency mapping after natural statistical method followed by feature extraction

First category is based on Google Nexus One and the other is based on Google Galaxy Nexus. The android device on which the tool is installed is connected to a desktop through USB cable which enables the tool usage. The software interface automatically selects one of the power models on the basis of device and enlists all the applications which are available on the device. The appropriate video application can be selected for the purpose of monitoring. The results are generated in terms of electric charge consumed by the application running on the device which makes the comparison more authentic.

5 Results

The proposed method is tested on 10 samples from different backgrounds like animated video, movie clips, and music videos. To make the approach more generic the duration of video samples and quality are kept different. The samples are encoded into the four standards: XVID, DIVX, MPEG and DIV3. The power is measured in milliampere-hours to check the effect in terms of electrical charge consumed by the sample in the device. The file size of modified samples is measured in Megabytes and is compared with the original one. Finally the usability rating of different samples is checked with 10 users, who have rated the same on the basis of individual experience. For analysis of usability, the mean value of user rating is taken for an individual sample with a particular codec. XVID, DIVX, MPEG and DIV3 are used for encoding the uncompressed sample. It is clear from the results that DIVX is the highest among all originally encoded samples in terms of power consumption followed by MPEG, DIV3 and XVID.

After finding the Region of Interest for the uncompressed samples, the features outside it are blurred for increasing redundancy at pixel level. These samples are again encoded for comparison with original one. The results in Table 1 infer that the order of power consumption in modified samples is same as in originally encoded samples. It can be noted that for different video codecs the percentage reduction in power is different, i.e. maximum of 7.6% and a minimum of 6.8%. This can be explained on the basis of fact that compression ratio for the different video codec is different. In addition to it, the video having a larger area outside Region of Interest will have relatively more redundant features which lead to higher compression.

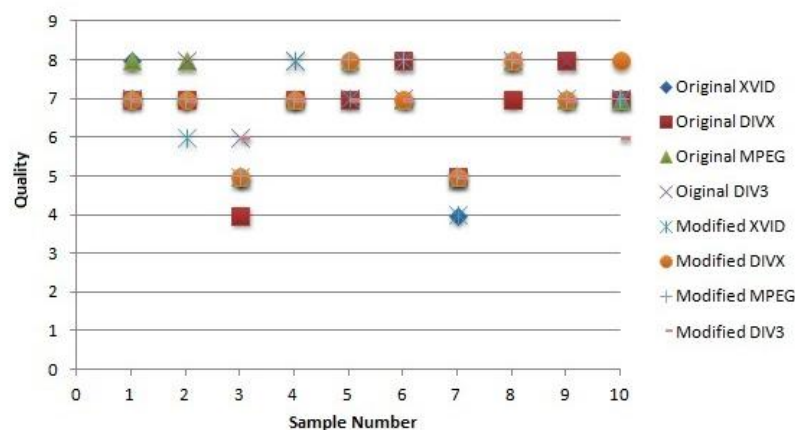
In Table 2 the different files are compared on the basis of their size. It is clear the DIVX and XVID have identical size in original as well as modified samples. The reduction in size varies from 24% to 46%. This is due to the fact that different video samples have different size of the region of interest and also different video codecs provide different compression. This leads to a weak correlation between size and power consumption on the basis of saliency mapping.

Table 1. Power consumed by original and modified video samples for four video codec “XVID, DIVX, MPEG and DIV3” in Milliamp Hours.

Sample	Duration	Power Consumption for Original Video Samples in mAh				Power Consumption for Modified Video Samples in mAh			
		XVID	DIVX	MPEG	DIV3	XVID	DIVX	MPEG	DIV3
1	5:16	8.96	9.04	8.99	8.97	8.34	8.4	8.37	8.35
2	3:03	5.33	5.37	5.35	5.34	4.96	4.98	4.96	4.97
3	7:57	13.6	13.69	13.67	13.66	12.6	12.71	12.67	12.71
4	5:00	8.16	8.22	8.19	8.17	7.56	7.6	7.59	7.57
5	4:03	6.96	6.99	6.98	6.97	6.46	6.47	6.47	6.44
6	7:23	12.33	12.41	12.37	12.35	11.44	11.49	11.46	11.45
7	4:23	7.45	7.49	7.48	7.46	6.92	6.95	6.94	6.93
8	5:06	8.62	8.66	8.64	8.63	8.0	8.05	8.03	8.02

Table 2. Size of original and modified video samples for four video codec “XVID, DIVX, MPEG and DIV3” in Megabytes.

Sample	Duration	Size of Original Video Samples in MB				Size of Modified Video Samples in MB			
		XVID	DIVX	MPEG	DIV3	XVID	DIVX	MPEG	DIV3
1	5:16	8.96	9.04	8.99	8.97	8.34	8.4	8.37	8.35
2	3:03	5.33	5.37	5.35	5.34	4.96	4.98	4.96	4.97
3	7:57	13.6	13.69	13.67	13.66	12.6	12.71	12.67	12.71
4	5:00	8.16	8.22	8.19	8.17	7.56	7.6	7.59	7.57
5	4:03	6.96	6.99	6.98	6.97	6.46	6.47	6.47	6.44
6	7:23	12.33	12.41	12.37	12.35	11.44	11.49	11.46	11.45
7	4:23	7.45	7.49	7.48	7.46	6.92	6.95	6.94	6.93
8	5:06	8.62	8.66	8.64	8.63	8.0	8.05	8.03	8.02

**Fig. 3.** Quality rating for all samples in different video codecs

The usability testing has been done over 10 people. The differently encoded samples are randomized and users have been asked to rate them in terms of quality in the scale of 10. The average round off value is taken for plotting the graph showing quality level.

It is clear from the graph shown in Fig. 3 that for a particular sample the user is not able to differentiate in terms of quality. In many places even the blurred video gets higher ratings from the original one.

6 Conclusion and Future Work

From the results obtained, it can be concluded that the possibility of saving power and storage space while maintaining an optimum usability level is very high. The power saving is up to 7.6 %, whereas file size is reduced by 46% without changing the hardware configuration for the codec. The effectiveness of the applied saliency mapping is responsible for maintaining quality of service (QoS). Users are not able to differentiate the quality as features outside the region of interest only are diminished. Since usability is a very subjective issue, accuracy of the saliency model plays an important role. This approach can be further extended to online streaming for graceful degradation in quality if bandwidth is limited.

References

- [1] A. Chimienti, L. Fanucci, R. Locatelli, S. Saponara. 2002." VLSI architecture for low-power video co-dec system". *Microelectronics Journal* 33 (5-6), 417–427. doi: 10.1016/S0026-2692(02)00009-5.
- [2] Atsushi Hatabu, Takashi Miyazaki And Ichiro Kuroda. 2002." QVGA/CIF Resolution MPEG-4 Video Codec Based on a Low-Power and General-Purpose DSP". Presented at Signal Processing Systems, San Diego, October 16-18.
- [3] B. D. Noble, M. Satyanarayanan, D. Narayanan, J. E. Tilton, and J. Flinn. 1997." Agile Application-Aware Adaptation for Mobility". *ACM SIGOPS Operating Systems* 31(5), 276-287. doi: 10.1145/269005.266708.
- [4] Chu-Hsing Lin, Jung-Chun Liu, Chun-Wei Liao. 2007. "Energy Analysis of Multimedia Video Decoding on Mobile Handheld Devices". Presented at Multimedia and Ubiquitous Engineering, Seoul, April 26-28.
- [5] Kihwan Choi , Dantu, K. , Wei-Chung Cheng and Pedram, M. 2002. "Frame-based dynamic voltage and frequency scaling for a MPEG decoder". Paper presented at IEEE Computer Aided Design, CA, USA, November 10-14.
- [6] Koen Danckaert, Kostas Masselos, Francky Catthoor and Hugo De Man. 1999." Strategy for power efficient combined task and data parallelism exploration illustrated on a QSDPCM video codec". *Journal of Systems Architecture* 45(10), 791–808. doi: 10.1016/S1383-7621(98)00039-3.
- [7] Krishna Rao Vijayanagar , Joohee Kim , Yunsik Lee and Jong-bok Kim. 2014. "Low complexity distributed video coding". *Journal of Visual Communication and Image Representation* 25(2), 361–372. doi: 10.1016/j.jvcir.2013.12.006.
- [8] Lingyun Zhang, Matthew H. Tong, GarrisonW. Cottrell. 2009. "SUNDAY: Saliency Using Natural Statistics for Dynamic Analysis of Scenes". Presented at Thirty-first Annual Cognitive Science Society Conference, Amsterdam, July 29- August 1.
- [9] Melloni L, van Leeuwen S, Alink A and Müller NG. 2012. "Interaction between bottom-up saliency and top-down control: how saliency maps are created in the human brain". *Cerebral cortex* 22 (12), 2943-2952. doi: 10.1093/cercor/bhr384.
- [10] Mohsen Jamali, Langroodi, Joseph Peters, and Shervin Shirmohammadi. 2014." Complexity Aware Encoding of the Motion Compensation Process of the H.264/AVC Video Coding Standard". Presented at Network and Operating System Support on Digital Audio and Video Workshop, New York, Singapore, March 19-21.
- [11] Morgan H. Miki, Gen Fujita, Takao Onoye, and Isao Shirakawa. 1997." Low-Power H .263 Video CoDec Dedicated to Molde Computing". Presented at Low Power Electronics and Design, Monterey, August 18-20.

- [12] Peter Longhurst, Kurt Debattista and Alan Chalmers. 2006." A GPU based Saliency Map for High-Fidelity Selective Rendering" Presented at 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa, Cape Town , January 25-27.
- [13] P. Ndjiki-Nya, D.Doshkov , H.Kaprykowsky , F.Zhang , D.Bull and T.Wiegand. 2012." Perception-oriented video coding based on image analysis and completion: A review". *Signal Processing: Image Communication* 27(6), 579–594. doi: 10.1016/j.image.2012.01.003.
- [14] Posner, M. I.1980. "Orienting of attention". *The Quarterly Journal of Experimental Psychology* 32, 3-25.doi: 10.1080/00335558008248231.
- [15] Praveen Kumar Yadav and N. Ramasubramanian. 2014. "Power Consumption of Android Device Using Different Video Codecs: An Analysis". Presented at Advance Computing Conference (IACC), Gurgaon, February 21-22.
- [16] Sebastian Montabone , Alvaro Soto. 2010. "Human detection using a mobile platform and novel features derived from a visual saliency mechanism". *Image and Vision Computing* 28, 391-402. doi: 10.1016/j.imavis.2009.06.006
- [17] Shao-Ping Lu and Song-Hai Zhang. 2011." Saliency-Based Fidelity Adaptation Preprocessing for Video Coding" *Journal of Computer Science and Technology* 26(1), 195-202. doi: 10.1007/s11390-011-9426-5.
- [18] S. Mohapatra and N. Venkatasubramanian. 2003. "PARM: Power-Aware Reconfigurable Middleware". Presented at 23rd International Conference on Distributed Computing Systems, Providence, May 19-22.
- [19] Surendar Chandra. 2003. "Wireless network interface energy consumption: implications for popular streaming formats", *Multimedia Systems* 9(2), 185 – 201. doi:10.1007/s00530-003-0089-0
- [20] T. Pering, Y. Agarwal, R. Gupta and R. Want. 2006." Coolspots: reducing the power consumption of wireless mobile devices with multiple radio interfaces". Presented at International Conference on Mobile Systems, Applications and Services, New York, June 19-22.
- [21] Viswanath Krishnamurthy, Swamy D. Ponpandi and Akhilesh Tyagi. 2011. "A novel thread scheduler design for polymorphic embedded systems". Presented at Compilers, Architectures and Synthesis for Embedded Systems (CASES), Taipei, October 9-14.
- [22] W. Barendregt and M. M. Bekker. 2006." Developing a coding scheme for detecting usability and fun problems in computer games for young children". *Behavior Research Methods* 38(3), 382-389. doi: 10.3758/BF03192791
- [23] Weiwei Fang , Yangchun Li , Huijing Zhang , Naixue Xiong , Junyu Lai and Athanasios V. Vasilakos, 2014. "On the throughput-energy tradeoff for data transmission between cloud and mobile devices". *Information Sciences* 283, 79–93. doi: 10.1016/j.ins.2014.06.022
- [24] Wei Song, Dian W. Tjondronegoro, Shu-Hsien Wang and Michael J. Docherty.2010." Impact of Zooming and Enhancing Region of Interests for Optimizing User Experience on Mobile Sports Video". Presented at 8th International Conference on Multimedia, Firenze, October 25-29.
- [25] Zhicheng Li, Shiyin Qin and Laurent Itti. 2011." Visual attention guided bit allocation in video compression". *Image and Vision Computing* 29(1), 1–14. doi: 10.1016/j.imavis.2010.07.001.
- [26] Zhihai He, Yongfang Liang, Ishfaq Ahmad, and Dapeng Wu. 2005. "Power-Rate-Distortion Analysis for Wireless Video Communication Under Energy Constraints". *Circuits and Systems for Video Technology* 15(5), 645 – 658. doi: 10.1109/TCSVT.2005.846433.