

Novel Intra-Prediction Framework for H.264 Video Compression using Decision and Prediction Mode

Pradeep Kumar N.S.

Associate Professor

Department of Electronics & Communication Engg.
South East Asian college of Engineering & Technology,
Bangalore, India

H.N. Suresh

Professor

Department of Electronics & Instrumentation Engg.
Bangalore Institute of Technology,
Bangalore, India

Abstract—With the increasing usage of multimedia contents and advancement of the communication devices (along with services), there is a heavy demand of an effective multimedia compression protocol. In this regards, H.264 has been proven to be an effective video compression standard; however, its computational complexity associated with out and various other issues has been impediment towards mainstream of research towards compression. Therefore, we present a novel framework that enhances the capability of H.264 compression method by emphasizing on accomplishing the cost effectiveness of computational operation during intra-prediction mode. A simple and novel encoding mechanism has been formulated using H.264/AVC using decision mode of macro block as well as selection of prediction mode exclusively for intra-prediction in H.264/AVC. The study outcome is found to offer a superior signal quality as compared to conventional H.264 encoding mechanism.

Keyword—Encoding Mechanism; H.264 / AVC; Intra-Prediction Mode; Video Compression; Visual Quality

I. INTRODUCTION

The joint development of ITU-T and MPEG has recommended H.264/AVC as an industry standard for video compression in the year 2003, which delivers 50% more compression efficiency as compared to previous standards as well as it supports very high video quality over channels of lower bit rates. This improved efficiency of H.264/AVC paves the foundation of many promising and potential applications and services on mobile devices over wireless networks with ease of coding, transmission and error resilience [1] [2] [3]. The gain in high performance on such devices with H.264/AVC is obtained at the very high cost of computation as the overhead of processing for power constraint mobile devices is a major tradeoff as more processing requires more power consumption. Therefore, it is an open research problem to conceptualize, design and device a mechanism involves into H.264/AVC encoding processes to achieve higher performance with least computational overhead by low complex implementations [4][5]. Typically, an H.264/AVC encoding process involves removal of spatial, temporal and statistical redundancy of video signal. The transformation of macro blocks, (a basic coding unit of 16 x 16 block of displayed pixel) by quantization of transform domain (spatial frequency components/ co-efficient) from spatial domain provides the considerable amount of

compression. The complexity of the process is reduced in the H.264/AVC as compared to other such processing or computation functions by means of efficient implementation of integer function and minimal transformation of the block size in this case. Modern applications which has been conceptualized in various domains of business applications such as collaborative communication, visual sensors based applications, advanced medical applications, entertainment, etc which runs on a synchronized heterogeneous devices and platform right from desktop, laptops, smart phones, custom devices till cloud commuting architecture extended to internet of things are image, multimedia and video. The successive growth of expectations of better Quality-of-Services (QoS) with general accepted visual quality in real time basis always poses a tradeoff between availability of bandwidth, storage, communication, computation and transmission [6][7][8]. This is a core reason for evolution of design of video compression as video codec. The futuristic aim is to cope up with the demand of rich visual dimension of video streaming application [9] [10]. Historically, the basic of the design of video codec were surrounded across a custom designed hardware to overcome the constraints of limited processing capacity by optimizing the computational capacity as an objective function. The inventions of the advanced processor have helped to overcome these issues and provided an improvement in performance in a reliable way as well as they were rate successful and had wider availability. As a result, the “software only approaches” of video codec become practicable [11] [12].

At present, there is much discussion about the trend of using H.264 compression standard owing its beneficial factors. There are also certain extents of research work being carried out towards overcoming the flaws in existing system. The proposed study presents one such work. Section A discusses about the existing literatures towards assessing the usage pattern of H.264 standard of multimedia compression. The existing research work in adoption of H.264 is discussed in Section II. Section III exhibits the problems being identified by the proposed study followed by brief highlight of the contribution of proposed system in Section IV. The algorithm being implemented in order to accomplish the research goal is discussed in Section V followed by discussion of result accomplished in Section VI. Finally Section VII gives some concluding remarks.

II. REVIEW OF LITERATURE

This section discusses about the existing research techniques towards adoption of H.264. We have reviewed about the existing techniques to improve the performance of video compression in our prior work [13]. This section of document closely investigates and writes an inference about intra-frame prediction and a reconstructing mode decision of video input signals. Our prior work has already reviewed existing literatures towards H.264 [13]. This section we further update about the related works. Qual et al. [14] in the year 2010 in their paper has classified fast intro predication mode decision into three categories based on block features, mode feature, and edge or directional information. Song et al. [15] have presented an approach of hierarchical intra-prediction method for video application for mobile based on orientation gradient based discretization of total variation. In their method, they shrink the candidate mode set in the Rate-Distortion Optimization (RDO) process. Lim et.al [16] introduces two mechanism 1) fast block size decision (FBS) and 2) fast mode decision (FMD) by considering a metric called similarity of the higher position pixels and left position pixels individually, which leads to a reduction of computational complexity. The simulation fallout shows maxima of 79% and 77% common instance investments with irrelevant cost in PSNR and bitrates. The coding performance of H.264/AVC can improve as per its operating conditions of types of modes as well the efficient optimization of the rate distortion in order to select the optimal mode. This approach introduces an additional overhead to the encoder as it requires compute Rate-Distortion (RD) price for the coding mechanisms. Lee et al. [17] have proposed a technique for quick mode assessment especially for the inter picture micro block (MB), which minimizes the cost for RDO. This method approaches to correlate spatial-temporal homogeneity for estimating the cost of the motion in both Intra and inter modes. A significant coding efficiency is maintained with relatively lower encoding time. Significance of edge feature is also laid by various research papers towards efficient coding scheme. In order to ensure optimal video compression technique, it is also essential to undergo the study of transcoding procedure to visualize the acceptability of the compatibility issues using H.264. Study in this direction was carried out by Liu et al. [18] where the authors have considered mobile transcoding of MPEG-2 video to H.264/AVC format. The outcome of the study shows that energy trend of DCT of macro-blocks of MPEG-2 is potentially correlated to the intra-prediction modes of H.264/AVC. Study towards fast intro-coding scheme was emphasized by Wu et al. [19], where the authors have discussed about the existing outcomes of the study and showed significant reduction in the encoding time, however, there exist low quality video while reconstruction. Similar work is presented by Shen et al. [20], where quad tree structured Coding Unit (QTS-CU) is used for provisioning the recursive splitting process into N- equal sized block, where N=4. The mechanism exhibits effective correlation among the three namely 1) prediction mode, 2) Motion Vector and 3) Rate distortion cost for varied depth levels and spatially-temporally coding units with overall minimization of 49% to 52 % computational complexities on multiple kinds of video sequences for different coding mechanisms. Studies toward

adoption of orthogonal modes elimination strategy was seen in the work of Peiman et al. [21]. The authors have used RD theory and selected only one of the orthogonal modes. Studies towards accelerating the mode decision process is also witnessed in literatures for the target of minimizing the quantity of methods mandatory to be tested for each macro blocks. For better coding efficiency, it is required to furnish better coding interoperability. Study in such direction was carried out by Su et al. [22] to provide the interoperability between MPEG-2 and H.264/AVC. The investigational outcomes disclose that typical 85% of computation time (a speed-up factor of seven) can be reduced compared to the encoding schemes. It is essential the video compression technique should also be tested on futuristic video file formats like UHD format. Studies in this direction has been carried by Lin et al. [23], in which the authors illustrates an interlaced block reordering scheme with a preliminary mode decision method to resolve the data dependency between intra-mode decision and reconstruction process in their encoding mechanism with 77% minimization of computational complexity, further more studies on algorithm efficiency on video compression was conceded by Lin et al. [23], where the authors proposes an efficient cascaded mode decision MPEG to H.264 I-frame encoding. The outcome of the study shows efficient PSNR, bitrates and reduced computation complexity performance. Therefore, there are various techniques towards encoding system presented in existing system.

III. PROBLEM IDENTIFICATION

The problem identified from the existing research techniques are highlighted as follows:

- Existing techniques directly implements H.264 without emphasizing over the possible computational complexity associated with it.
- Majority of the present approaches doesn't emphasize on the encoding process keeping in mind about the low powered devices with limited computational capability.
- The compression algorithm implemented till date using encoding techniques incorporates sophisticated mode decision approaches that may result in significant degradation of visual quality of the transmitted signal.
- The existing approaches using intra-prediction mode doesn't emphasize over the potential computational complexity associated with deblocking filters.
- The mechanism of minimizing the number of modes for assessing the performance of rate distortion theory was less attended in existing literatures by the researchers.

Therefore, it can be seen that along with contribution, the existing literatures have significant amount of problems that are open ended yet. The next section briefs about proposed solution.

IV. PROPOSED SOLUTION

The prime aim of this paper is to evolve up with a framework that uses H.264 for taking decision of macro-block mode along with selection of mode of prediction. The sole

purpose is to perform improvement over intra-prediction deploying H.264 standard. The architectural scheme is shown in Fig.1.

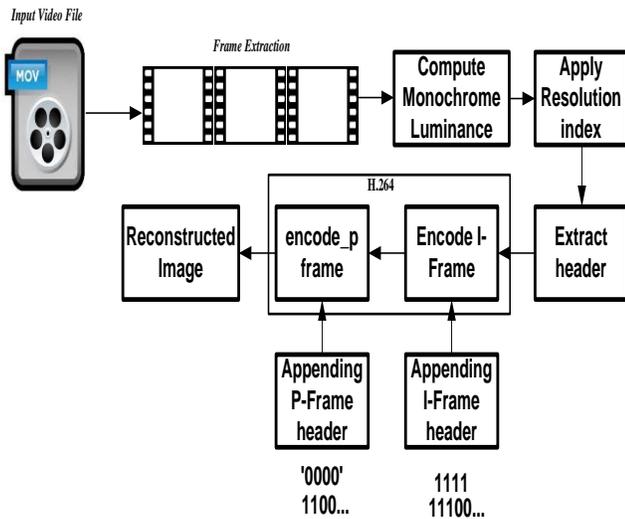


Fig. 1. Architectural Scheme of Proposed System

The proposed system takes the input of video resulting in conversion of frames which is further subjected to computation of monochrome luminance for 4x4, 8x8, as well as 16x16 blocks as intra-prediction modes over vertical direction horizontal direction, DC, as well as plane mode. Different from any existing prediction technique, the proposed system computes cost of rate distortion by applying resolution index. Header information is extracted height, weight, quantization parameter, size of block, and frame ranges followed by encoding of I and P frame sequentially to obtain reconstructed image. The next section outlines the algorithm implemented for this purpose.

V. ALGORITHM IMPLEMENTATION

This proposed system offers a mechanism of encoding and decoding of specific frame using H.264 encoder in order to improve the prediction selection mode. The algorithm takes the input of the video file converts it into frame, which after being subjected to novel enhanced H.264 encoder yields an output of reconstructed signal. The steps involved in the proposed algorithm are as follows:

Video Compression technique using H.264 codec

Input: I_{frames} , F_{rang} , QP , i, j, k , F_{size} , IB_{size}

Output: $I_{recframes}$

Start

1. $I = [I_{frames}]$
2. $R \leftarrow \text{arb}(F_{rang}, F_{size})$
3. $Y \leftarrow 255 * [\text{convert RGB to Gray} (I, R)]$
4. $K \leftarrow [F_{size}, IB_{size}, QP]$
5. $Y \leftarrow [Y, H.264_{Encode}, K]$
6. for $K=1:j$

7. $I_{rec} \leftarrow H.264_{Decode}(Y, R, i)$

8. End

9. $I_{recframes} \leftarrow \text{rec} [I_{rec}, I_{frame}]$

End

The flow of the proposed algorithm is as follows: The algorithm initially reads the complete input video and extracts the frames based on configured start and end frame values. This step is further followed by implementation of H.264 codec with feature to opt for specific quantization parameter with specific selection of multiple frame size e.g. i) QCIF (144 x 176), ii) CIF (288 x 352), iii) WVGA (480 x 800), and iv) HD (720 x 1280). The algorithm also considers intra-block size of 4x4, 8x8, and 16x16. The selected frames of the video are indexed and are subjected to calculation of monochrome luminance. The algorithm also extracts the respective information about the height and width of the frame and extracts header information from the respective sample. A discrete function is made for the header file which considers the following input arguments e.g. height, width, quantization parameter, frame range, and size of block etc. This is basically the mechanism of performing the encoding of I frame that is further followed up by the encoding of the P-frame. For the purpose of encoding, we apply 4 times the size of the block when we perform encoding of P-frames. While performing encoding of I frame, we append 1111 to the I-frame header and consider it to be a bit stream. Similarly, the study considers appending 0000 as a part of appending P frame header in order to generate the bit stream. The encoding of the p frame is carried out considering an array mainly consists of frame information and quantization parameter that after processing yields the output of encoded P frame. The dependable parameters used for this encoding techniques is lagranian multiplier used as i) an essential component to perform motion estimation and ii) in selection of partition mode.

This encoding mechanism is carried out using a novel prediction-based strategy that implements a motion prediction of non-translation origin. This technique basically adopts a mechanism where two dimensional functions is developed using elastic motion model in order to perform evaluation of motion of non-translational origin existing among the structured blocks. The proposed algorithm implements a mode decision system whereby the variance score existing among the macro-blocks are utilized for performing an effective modeling of intra frame prediction mode decision. The algorithm considers an assumption that complexity associated with the texture basically corresponds to variance of the macro blocks. The operation of the algorithm is basically classified into two stages wherein the first stage corresponds to the computation of variance factor among the macro blocks that are required to be encoded as well as the threshold factor too. However, if the variance of the macro blocks is found to be more than the threshold factor that the mechanism selects only 14-18 macro blocks or else it performs selection of more number of macro blocks greater than 18. This mechanism is found to offer faster response time as compared to the existing algorithms.

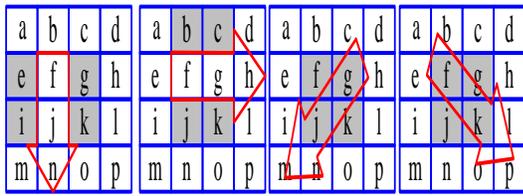


Fig. 2. Analysis of Multiple Orientations

Fig.2 shows the multiple forms of orientation in the order to vertical orientation, horizontal orientation, diagonal down-left orientation, and diagonal downright orientation. The algorithm performs respective computation of intensity gradient over all the direction followed by selection of significant three modes that has minimal score of gradient intensity of the pixel. In this case, the approach also assumes DC and most probable mode have higher probability of selection as modes for candidate prediction. The computation of the rate distortion is then carried out from the finally chosen modes of candidate prediction. At the same time, the encoding of macro blocks is carried out using H.264 over the mode that has been explored to posed minimal cost of rate distortion. The study considers a prediction mode to be equivalent to most probable mode that is assessed to have minimal mode of prediction quantity between the blocks of top neighbor and left neighbor. On the other hand, the system considers switching over to DC mode from most probable mode in case there is no availability of top neighbor block and left neighbor block. At the same time, it is highly feasible to have significant functions for gradient intensities with equivalent score of gradient intensities as it is believed to possess integer precision only. In an adverse scenario, the number of the gradient intensities possessing similar number of gradient could have only 8 as a value. Therefore, in such cases, the selection of the modes for candidate prediction is carried out on the basis of the modes that are found with minimized number of prediction among the gradient intensity. The prime reason behind this is the arrangement of the prediction number that is carried out using frequencies. On the other hand, if the algorithm chooses the vase of 18 macro blocks than the mode of selection prediction is carried out by dividing the 8x8 blocks into 2x2 sub-blocks of 16 numbers. This is carried out as shown in Fig.3. Finally, all the end sub-blocks are subjected to averaging all the sub-blocks obtained. a_8

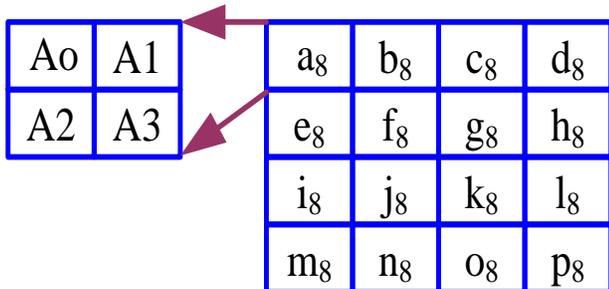


Fig. 3. 4x4 Blocks from 8x8 Blocks

Interesting fact to notice here is that selection of the 14 macro blocks prediction is found nearly same as the next process. However, if the system finds less number of gradient intensity value that is found to be equivalent to most probable mode than the system fairly implements the selection strategy for gradient intensity with two modes only in the form of candidate mode. This step potentially minimizes the computational complexity associated with encoding multimedia video of larger dimension or resolution. Finally, in case of 116 macro blocks, we divide the macro blocks in 4x4 blocks of 16 numbers. All the pixels associated with sub blocks are averaged followed by computation of gradient orientation. In this phase, the selection of the modes of candidate prediction is carried out by on the basis of the minimal value of modes with gradient intensity and DC. For the purpose of computing the cost of rate distortion, we perform selection of prediction modes only. Therefore, the proposed system maintains a good balance between the encoding operation as well as computational complexity by introducing a novel mode decision based on macro blocks as well as selection of prediction mode. The next section discusses about the outcome being accomplished in the proposed system.

VI. RESULT ANALYSIS

The implementation of the proposed study was carried out in Matlab considering the video dataset from [24]. The prime observation being carried out over the encoding implementation work is mainly the size of the video, number of original bits, numbers of compressed bits, compression ratio, and Peak Signal-To-Noise Ratio (PSNR).



Fig. 4. Original Input frame



Fig. 5. Resized Luminosity



Fig. 6. Reconstructed Frames

The study outcome was testified using multiple forms of performance parameters exhibited in Table 1. The greater the PSNR, the improved the excellence of the compressed or reconstructed image. The MSE and PSNR is the two error matrix utilized to look at picture compression quality. MSE

signifies the increasing squared error amongst the compressed and the unique image, while PSNR signifies an amount of peak errors present in an image. The lesser the value of MSE, lower the pear errors. To calculate the PSNR, the primary block is initially calculated utilizing the MSE with following equations:

$$PSNR(dB) = 10 * \log \left(\frac{255^2}{MSE} \right)$$

$$MSE = \sum_{i=1}^x \sum_{j=1}^y \frac{(A_{ij} - B_{ij})^2}{x * y}$$

TABLE. I. COMPARATIVE PERFORMANCE ANALYSIS

	Video-1		Video-2		Video-3	
	Proposed H.264	Existing H.264	Proposed H.264	Existing H.264	Proposed H.264	Existing H.264
Video Size	QCIF (144 X 176)		CIF (288 X 352)		WVGA (512 X 832)	
Intra Block Size	4	4	8	8	16	16
QP	16	16	16	16	16	16
No. of Original Bits	2027520	2027520	8110080	8110080	34078720	34078720
No. of Compressed Bits	1166726	629941	2319575	1116281	32823374	23971812
Encoding Time (Sec)	21.400980	27.65221	65.189897	872.66512	101.196864	127.5429
Mean PSNR	47.8189	23.6518	52.1431	31.2887	51.9164	38.8721

The study outcome has been testified using the proposed mechanism of encoding using H.264 and conventional mechanism of H.264. We find that in every sense the performance of the proposed system has excelled better as compared to existing H.264. We find that proposed system offer better compression performance that is highly suitable for transmission of multimedia contents over wireless channel. The lower outcome of encoding time also elaborates the fact that proposed system offer better solution to computational complexity by offering faster response time. Finally, the proposed system offer better signal quality as seen from the PSNR of the reconstructed signal.

VII. CONCLUSION

The projected scheme has emphasized an exclusive video compression and reconstruction scheme for different video formats. In all the cases presented in this paper, the Mean PSNR of the decoded video sequence always exhibit a similar behaviour; i.e. the quality steadily decreases with the

increment of the quantization factor. The results obtained in these experiments opinions out the great association among the AQI and the human visual system for H.264 video coded sequences, in contrast with the Mean PSNR, as a reliable way to measure the perceptual quality of images. This fact opens the possibility of incorporating self-regulated compression parameters depending on the perceptual quality. It also gives good compression ratio as well as efficient encoding timer is used to encode the videos. PSNR is very commonly used in measuring the encoded video quality, but the drawback is PSNR is not totally correlated to the subjective quality of the video. This means a human being may feel a lower PSNR video has better quality than a higher PSNR video which is compressed from the same video sequence.

REFERENCES

- [1] I. E. Richardson, "The H.264 Advanced Video Compression Standard", John Wiley & Sons, 2011
- [2] B. Juurlink, M. A-Mesa, C. C. Chi, A. Azevedo, C. Meenderinck, A. Ramirez, "Scalable Parallel Programming Applied to H.264/AVC

- Decoding”, Springer Science & Business Media-Technology & Engineering, 2012
- [3] X. Tian, T. M. Le, Y. Lian, “Entropy Coders of the H.264/AVC Standard: Algorithms and VLSI Architectures”, Springer Science & Business Media-Technology & Engineering, 2010
- [4] W. Jackson, “Digital Video Editing Fundamentals”, APRESS, 2016
- [5] X. Tian, T. M. Le, Y. Lian, “Entropy Coders of The H.264/AVC Standard: Algorithms And VLSI Architectures”, Springer Science & Business Media, 2010
- [6] Y. Zhang, Z. Han, “Contract Theory for Wireless Networks”, Springer, 2017
- [7] K. A-Begain, A. Ali, “Multimedia Services And Applications In Mission Critical Communication Systems”, IGI Global, 2017
- [8] W-C. Hu, “Multidisciplinary Perspectives On Telecommunications”, Wireless Systems, And Mobile Computing, IGI Global, 2013
- [9] M. E. Vermaat, S. L. Sebok, S. M. Freund, M. Frydenberg, J. T. Campbell, “Enhanced Discovering Computers”, Cengage Learning, 2016
- [10] R. Trestian, “Convergence of Broadband, Broadcast, And Cellular Network Technologies”, IGI Global, 2014
- [11] K. K-S, S. Ready, “Practical Guide to Machine Vision Software: An Introduction With Labview”, John Wiley & Sons, 2015
- [12] J. J. Parsons, D. Oja, “New Perspectives on Computer Concepts”, Comprehensive, Cengage Learning, 2014
- [13] P. Kumar N.S., H.N. Suresh, “Studying an Effective Contribution of Techniques of Video Compression”, International Journal of Electronics Communication and Computer Engineering, Vol. 6, Issue.1, 2015
- [14] D. Quan, Y-S Ho, "Categorization for fast intra prediction mode decision in H.264/AVC," IEEE Transactions on Consumer Electronics, Vol.56, No.2, pp.1049, 1056, May 2010
- [15] Y. Song, J. Long, K. Yang, G. Yang, "Complexity scalable intra-prediction mode decision algorithm for mobile video applications," IEEE-IET Communications, Vol.8, No.9, pp.1654-1662, 2014
- [16] K. Lim, S. Kim, J. Lee, D. Pak, S. Lee, "Fast block size and mode decision algorithm for intra prediction in H.264/AVC", IEEE Transactions on Consumer Electronics, Vol.58, No.2, pp.654-660, 2012
- [17] J. Y. Lee, H. W. Park, "A Fast Mode Decision Method Based on Motion Cost and Intra Prediction Cost for H.264/AVC", IEEE Transactions on Circuits and Systems for Video Technology, Vol.22, No.3, pp.393-402, 2012
- [18] X. Liu, K-Y Yoo, S. W. Kim, "Low complexity intra prediction algorithm for MPEG-2 to H.264/AVC transcoder," IEEE Transactions on Consumer Electronics, Vol.56, No.2, pp.987-994, 2010
- [19] C-Y Wu, P-C Su, "Fast Intra-Coding for H.264/AVC by Using Projection-Based Predicted Block Residuals," IEEE Transactions on Multimedia, Vol.15, No.5, pp.1083-1093, 2013
- [20] L. Shen, Z. Zhang, Z Liu, "Adaptive Inter-Mode Decision for HEVC Jointly Utilizing Inter-Level and Spatiotemporal Correlations", IEEE Transactions on Circuits and Systems for Video Technology, Vol.24, No.10, pp.1709,1722, 2014.
- [21] H. Pejman, F. Zargari, "An efficient fast intra mode decision method based on orthogonal modes elimination," IEEE Transactions on Consumer Electronics, Vol.58, No.4, pp.1445-1452, 2012
- [22] C. Su, Y. Lin, "Zero-block inter/intra mode decision for mpeg-2 to H.264/AVC inter P-frame transcoding," IEEE-IET Image Processing, Vol.4, No.6, pp.494-504, 2010
- [23] Y. Lin, J-D Wu, "Cascaded mode decision for MPEG-2 to H.264/AVC intra frame transcoding," IEEE Transactions on Consumer Electronics, Vol.55, No.4, pp.2172, 2177, 2009
- [24] "Database: Image & Video Clips (2)", http://see.xidian.edu.cn/vipsl/database_Video.html, Retrieved, 15th March, 2017