264/AVC VIDEO ENCODER FOR FAST INTER-PREDICTION MODE DECISION

[1] S. RAGHU VINODH
[2]DR. PANDARINATH POTLURI
[1]RESEARCH SCHOLAR [2]PROFESSOR
[1]CMJ UNIVERSITY [2]GUDLAVALLERU ENGINEERING

ABSTRACT

H.264/MPEG-4 Part 10 or AVC (advanced video coding) is currently one of the most widely used industry standards for video compression. There are several video codec solutions, both software and hardware, available in the market for H.264. This video compression technology is primarily used in applications such as video conferencing, mobile TV, blu-ray discs, digital television and internet video streaming.

In real time applications such as video streaming and video conferencing it is important that the video encoding/decoding is fast. It is known, that most of the complexity lies in the H.264 encoder, specifically the motion estimation (ME) and mode decision process introduces high computational complexity and takes a lot of CPU (central processing unit) usage. The mode decision process is complex because of variable block sizes (16X16 to 4x4) motion estimation and half and quarter pixel motion compensations.

Innovations in the communication systems have been tremendous in the last decade. Technology in communication systems has transformed from having only analog television via cable, satellite with

© 2013 IJAIR. ALL RIGHTS RESERVED

availability of only few channels or mobile phones that can only make voice calls or send SMS messages or internet connections that are slow, mostly connected through dial –up modem connected via telephone lines, back then data was stored on floppy disks, magnetic tapes and bulky hard drives.

Today the world has transformed into the so called “digital age” or “electronic age”, where mobile phones are called smart phones because they can not only make phone calls but are also used for web browsing, sending emails, watching / capturing videos , transfer data, navigation purposes and as camera. Digital television sets have become more compact with availability of regional and international channels with HD quality. Data is stored on re-writable DVDs, Blu-ray discs and hard disks which are light weight, portable with huge space for storage. Internet connection is blazing fast with wireless routers and modems operating at faster speeds [4]. In this fast growing world of communications, data compression is still one of the most essential components in any multimedia system. Modern data compression techniques offer the possibility to store or transmit the vast amounts of data necessary to represent digital videos and images in an efficient and robust way.

A detailed experimental study of face detection algorithms based on “Skin Color” was read. Three color spaces, RGB, YCbCr and HSI are of main concern. They compared the algorithms based on these color spaces and have combined them to get a new skin-color based face-detection algorithm that improves accuracy. Experimental results show that the

© 2013 IJAIR. ALL RIGHTS RESERVED

proposed algorithm is good enough to localize a human face in an image with an accuracy of 95.18% [2].

Another face detection algorithm uses color images in the presence of varying lighting conditions as well as complex backgrounds. The method detects skin regions over the entire image, and then generates face candidates based on the spatial arrangement of these skin patches. The algorithm constructs eye, mouth, and boundary by using a transfer of color space from RGB to YCbCr maps for verifying each face candidate.

Template matching algorithms

Cross correlation is a template matching algorithm that estimates the correlation between two shapes that have a similar orientation and scale. Consider two series x(i) and y(i) where i=0,1,2...N-1. The cross correlation r at delay d is defined as

$$r = \frac{\sum_i [(x(i) - mx) * (y(i-d) - my)]}{\sqrt{\sum_i (x(i) - mx)^2} \sqrt{\sum_i (y(i-d) - my)^2}}$$

Where mx and my are the means of the corresponding series. If the above is computed for all delays d=0, 1, 2,.. N-1 then it results in a cross correlation series of twice the length as the original series.

Image based algorithms

- Statistical approach
- Neural networks[4]

Many commercial applications of face recognition are also available such as security system, criminal identification, and film processing. Like face detection face recognition

© 2013 IJAIR. ALL RIGHTS RESERVED

can also be categorized into tree types. They are

- Feature-based approach,
- Holistic approach and
- Hybrid approach.

<u>Packetization</u>

The first step in the process of multiplexing is packetization. This refers to formatting the long stream of data into blocks called packets. Instead of transmitting the data as a series of bytes, when formatted into blocks, the network can transmit a long stretch of data more reliably and efficiently. A packet mainly consists of two parts. First one is the header which contains the information about the data that it is carrying followed by the payload, which is the actual data.

In the application under consideration, the data that needs to be packetized are the audio and video streams. In case of transmitting more than one program, we have many video and audio streams. So, during the packetization, adopted method should be such that it would enable the user to easily realign the packets, at the de-multiplexer side, to form the corresponding streams. In order to ensure the above mentioned criteria and to meet the transmission channel requirements two layers of packetization are carried out. The first layer of packetization yields the packetized elementary stream (PES) and the second layer yields the transport stream (TS). This second layer is what is used for transmission. This process is shown in Fig. 4.2. Multiplexing takes place after the second layer of packetization, just before the transmission.

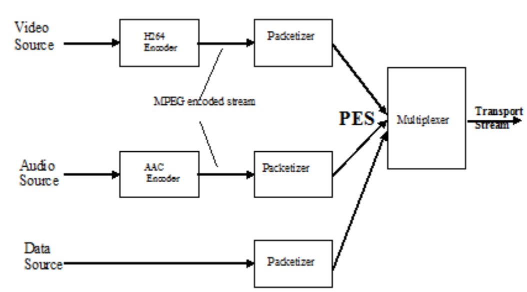

Fig.: Two layers of packetization

© 2013 IJAIR. ALL RIGHTS RESERVED

H.264 Encoder:

H.264 encoder works on the same principles as that of any other codec. Figure 2.6 shows the basic building blocks of H.264 video codec.

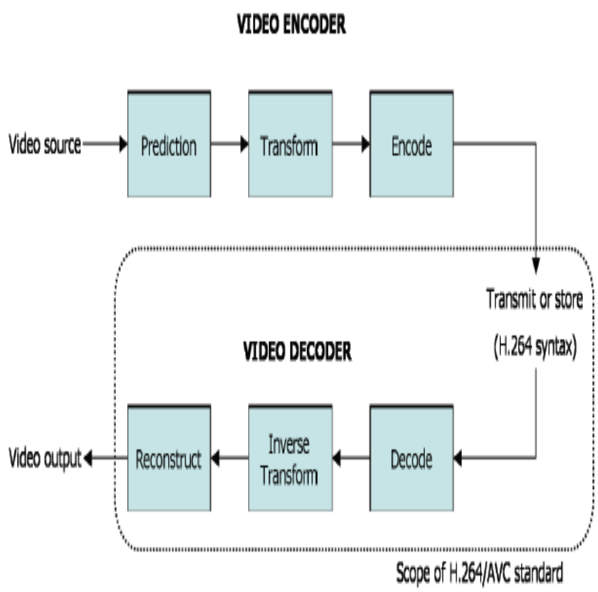

Figure 2.6 H.264 video coding and decoding process [4].

The input to the encoder is generally an intermediate format stream, which goes through the prediction block; the prediction block will perform intra and inter prediction (motion estimation and compensation) and exploit the redundancies that exist within the frame and between successive frames. The output of the prediction block is then transformed and quantized. An integer approximate of the discrete cosine transform is used (DCT) for transformation [12]. It uses 4x4 or 8x8 integer transform, and outputs a set of coefficients each of which is a weighted value for a standard basis pattern. The coefficients are then quantized i.e. each coefficient is divided by an integer value. Quantization reduces the precision of the transform coefficients according to the quantization parameter (QP). Typically, the result is a block in which most or all of the coefficients are zero, with a few non-zero coefficients. Next, the coefficients are encoded into a bit stream.

Conclusions:

From the simulation results described in Chapter 4, it can be concluded that the proposed fast adaptive termination algorithm based JM encoder is faster than the JM reference software. This is because when the JM reference software uses rate distortion optimization (RDO) which examines all possible combinations of coding modes, unlike the complexity reduction algorithm which makes use

© 2013 IJAIR. ALL RIGHTS RESERVED

of early skip mode detection, adaptive thresholds and homogeneity detection.

From Tables 4.1 through 4.4, it can be observed that there is an average of 43.20% reduction in encoding time in case of fast adaptive termination based encoder with negligible loss of PSNR and SSIM and a tolerable increase in bit-rate.

Figures 4.5 through 4.11, show plots of comparison of encoding time, PSNR, bit-rate and SSIM between the JM reference software and fast adaptive termination based JM encoder. The simulation was performed on CIF and QCIF sequences. Figures 4.12 through 4.15 show plots of comparison of PSNR, bit-rate, encoding time and SSIM between JM reference software and fast adaptive termination based JM encoder, using coastguard_qcif sequence at QP = 27. The simulation was performed for various values of QP = 22, 27, 32, 37. These simulation results again concur that, significant encoding time can be reduced by using the proposed complexity reduction algorithm and at the same time, fidelity of the input video stream is maintained.

References :

1. Open source article, "Intra frame coding" : http://www.cs.cf.ac.uk/Dave/Multimedia/node248.html

2. Open source article, "MPEG 4 new compression techniques" : http://www.meabi.com/wp-content/uploads/2010/11/21.jpg

3. Open source article, "H.264/MPEG-4 AVC," Wikipedia Foundation, http://en.wikipedia.org/wiki/H.264/MPEG-4_AVC

© 2013 IJAIR. ALL RIGHTS RESERVED

4. I.E.Richardson,"The H.264 advanced video compression standard",2nd Edition ,Wiley 2010.

5. R. Schafer and T.Sikora,"Digital video coding standards and their role in video communications,"Proceddings of the IEEE Vol 83,pp. 907-923,Jan 1995.

6. G.Escribano et.al,"Video encoding and transcoding using machine learning," MDM/KDD'08,August 24,2008,Las Vegas,NV,USA.

7. D. Marpe, T. Wiegand and S. Gordon, "H.264/MPEG4-AVC Fidelity Range Extensions: Tools, Profiles, Performance, and Application Areas", Proceedings of the IEEE International Conference on Image Processing 2005, vol. 1, pp. 593 - 596, Sept. 2005.

8. ITU-T Recommendation H.264-Advanced Video Coding for Generic Audio-Visual services.

© 2013 IJAIR. ALL RIGHTS RESERVED