

A Novel Approach towards Video Compression for Mobile Internet using Transform Domain Technique

Dhaval R. Bhojani

Research Scholar, Shri JIT University, Jhunjunu,
Rajasthan, India

Ved Vyas Dwivedi, PhD.

Director & Principal
Noble Group of Institutes Junagadh, Gujarat, India

ABSTRACT

As the era, in the internet applications, is of mobile internet, the use of the internet on the cell-phone networks and other mobile networks is increased so much. Also the data rates available on any mobile networks are limited so that the multimedia file transfers on the mobile networks require more time and cost. To reduce the data rates required to transfer multimedia files on the mobile networks, we require one compression algorithm which gives us higher compression with easiness of implementation. So here the paper present one algorithm which is much simpler to implement, as it is based on MPEG-2 then also gives compression in considerable amount, which is comparable to MPEG-4 techniques. The algorithm is implemented on MATLAB version 7.10.0.499a platform. The input to the encoder is in uncompressed '.avi' format. The testing of the compressed videos is done using MSU video quality measurement tool.

General Terms

Video compression algorithm useful for mobile internet applications

Keywords

Adaptive Rood Pattern Search (ARPS), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Group of Pictures (GOP), Run Length Encoding (RLE).

1. INTRODUCTION

As the usage of internet connectivity is increasing day by day on mobile networks, the need of video compression algorithm increases for the better compression ratio. Simultaneously the algorithm is having requirement of lesser complexity of design and easiness for implementation, as the mobile gadgets are having limitations of their own from complexity and prize point of view.

Now a day the video compression algorithm which is widely used for TV broadcasting is MPEG-4. This gives us the compression ratio near about 95%. But the complexity of the encoders and decoders which are used for the algorithm is such high, as it is object based compression algorithm [1]. Also it requires the processor which is much faster in processing than used for mobile gadgets.

So here, the authors of paper, concentrated on the much simpler algorithm to implement on relatively slower processors of mobile gadgets. Here authors are modifying the MPEG-2 algorithm, which is based on subjective compression, from the transform domain perspective, which is used in the intra frame compression of the video. Here paper introduces a new approach for converting the frames in frequency domain in such a way that using MPEG-2 we can achieve the compression ratio approximately equal to MPEG-4, with lesser complexity of the encoder and decoder.

The section-2 of the paper explains MPEG-2 in detail with some theoretical aspects also. Then section-3 explains the modification done in the encoder and decoder of MPEG-2. Section-4 of paper theoretically explains the parameters used to measure quality of video. Then section-5 shows the comparative results for both techniques, and finally section-6 concludes the paper.

2. REGULAR MPEG-2 ALGORITHM

The basic video compression system comprises of the video encoder at the transmitter, which encodes the video to be transmitted in terms of bits and the video decoder at the receiver, which reconstructs the video in its original form from the bit sequence received. The sub systems of encoder and decoder are discussed here for theoretical understandings.

2.1 Encoder

The video encoder for MPEG-2 system is as shown in Fig. 1:

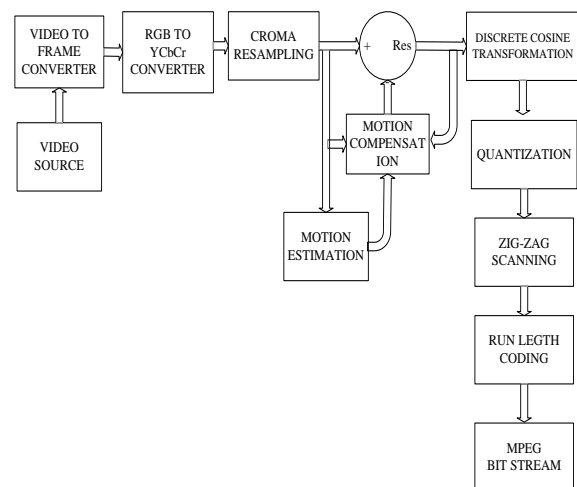


Fig. 1: MPEG-2 video encoder

In the encoder of regular MPEG first of all, the video is converted into the sequence of frames which are nothing but like still images sequence. Then the 10 sequential frames are selected (GOP) and format of the sequence is changed from RGB to YCbCr. Third stage re-samples the chrominance components of the frames from 4:4:4 to 4:2:0, as human eyes are less sensitive towards the chrominance components we can reduce no. of samples in that for compression. Then in group of pictures (GOP) the sequence is converted into sequence of IBBPBBPBB with use of motion estimation and compensation technique.



Fig. 2: Group of Pictures [2,3]

The next step is to take DCT of the motion compensated frames, which gives us coefficients matrix of same size as frame. The coefficients of the DCT of image are quantized using the standard quantization matrix.

The quantized coefficients of the transformed image are scanned in zig-zag manner. The encoding process of these coefficients are done using run length coding for the better compression and the bit stream of the coder is stored on any medium or transmitted. At last the bit-stream is formed and transmitted.

The theoretical concepts related to above encoder are described briefly here as below:

The first of several steps in the compression is to translate red, green and blue intensity information in each pixel into luminance/brightness values (Y), as well color vectors (Cb, Cr). The chrominance information can then be sub sampled.

There are three formats [2]:

- **4:4:4**

The chrominance and luminance planes are sampled at the same resolution.

- **4:2:2**

The chrominance planes are sub sampled at half resolution in the horizontal direction.

- **4:2:0**

The chrominance information is sub sampled at half the rate both vertically and horizontally.

The consecutive frames are correlated with each other so we can predict one from other and getting residue from them which will give us temporal compression in the video. So according to this there are three kinds of frames in GOP: I-frame, P-frame and B-frame.

An I-frame, or intra frame, is a self-contained frame that can be independently decoded without any reference to other images. The first image in a video sequence is always an I-frame. I-frames are needed as starting points for new viewers or resynchronization points if the transmitted bit stream is damaged. I-frames can be used to implement fast-forward, rewind and other random access functions. An encoder will automatically insert I-frames at regular intervals or on demand if new clients are expected to join in viewing a stream. The drawback of I-frames is that they consume much more bits, but on the other hand, they do not generate many artifacts.

A P-frame, which stands for predictive inter frame, makes references to parts of earlier I and/or P frame(s) to code the frame. P-frames usually require fewer bits than I-frames, but a drawback is that they are very sensitive to transmission errors because of the complex dependency on earlier P and I reference frames [3].

A B-frame, or bi-predictive inter frame, is a frame that makes references to both an earlier reference frame and a future frame.

Motion-compensated prediction is a powerful tool to reduce temporal redundancies between frames and is used extensively in MPEG-1 and MPEG-2 video-coding standards as a prediction technique for temporal DPCM coding. The concept of motion compensation is based on the estimation of motion between video frames, i.e., if all elements in a video scene are approximately spatially displaced, the motion between frames can be described by a limited number of motion parameters (i.e., by motion vectors for translatory motion of pixels). In this simple example the best prediction of an actual pixel is given by a motion-compensated prediction pixel from a previously coded frame. Usually both prediction error and motion vectors are transmitted to the receiver. However, encoding one motion information with each coded image pixel is generally neither desirable nor necessary. Since the spatial correlation between motion vectors is often high, it is sometimes assumed that one motion vector is representative for the motion of a "block" of adjacent pixels. To this aim, images are usually separated into disjoint blocks of pixels and only one motion vector is estimated, coded, and transmitted for each of these blocks.

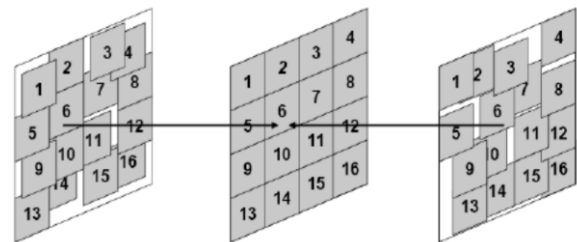


Fig. 3 Motion Estimation [11]

In the MPEG compression algorithms the motion compensated prediction techniques are used for reducing temporal redundancies between frames and only the prediction error images the difference between original images and motion compensated prediction images are encoded. In general, the correlation between pixels in the motion compensated inter frame error images to be coded is reduced compared to the correlation properties of intra frames in Fig. 3 due to the prediction based on the previous coded frame.

The motion compensated frames are ready for transmission now. So to convert it in frequency domain normal MPEG takes DCT of the frames, which is similar to the more familiar Fourier transform. This yields a series of coefficients indicating the magnitude of cosine functions at increasing frequencies so that the original signal can be reassembled in the spatial domain.

The quantization step truncates some of the least significant bits of information, making some coefficients go to zero. These coefficients are then zig-zag scanned and then entropy coded, to convert the coefficients into variable bit-length codes with the most common coefficients being coded with the fewest number of bits. This coding scheme is sometimes called Run length coding of Variable length coding.

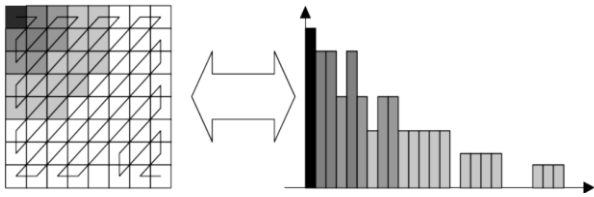


Fig. 4 Zig-Zag Scanning [4]

This kind of compression only removes redundant information within one frame, and one of the properties of video signals is that there is much redundant information repeated from frame to frame. More compression can be achieved by not re-transmitting these static portions of the picture.

2.2 Decoder

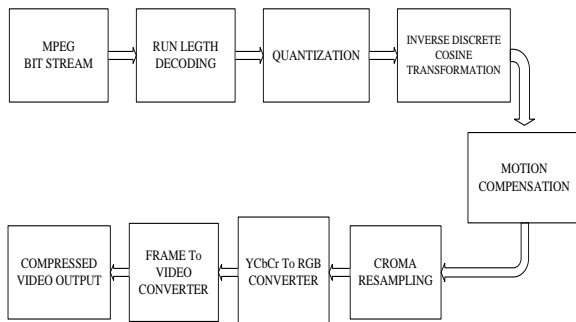


Fig. 5: MPEG-2 video decoder

First of all, as the receiver receives the MPEG bit stream the data bits are run length decoded and the quantized coefficients of DCT are achieved. Then these coefficients are de-quantized using the same quantization matrix and inverse DCT is taken of the coefficients, which gives us the motion compensated frames. Then from these GOP the original frames of video are achieved using motion compensation and estimation algorithms. Then chrominance components are re-sampled and frames are converted back to RGB format from YCbCr format.

3. MODIFIED MPEG-2 ALGORITHM

3.1 Encoder

The video encoder which has been implemented is as shown in Fig. 1:

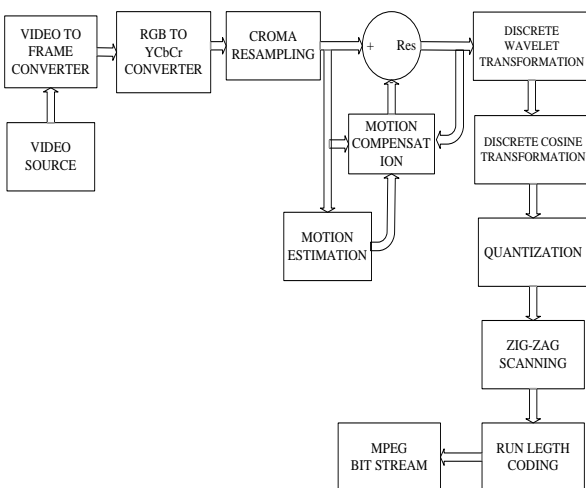


Fig. 6: Modified MPEG-2 video encoder

The modification is done in the transformation step of the encoder. As the motion compensated frames are available, the next step is to take DWT of the motion compensated frames, which gives us four components namely: Approximate, Horizontal, Vertical and Diagonal. Out of these four components only approximate component is transmitted so DCT of it is taken. The coefficients of the DCT of image are quantized using the standard quantization matrix.

The quantized coefficients of the transformed image are scanned in zig-zag manner. The encoding process of these coefficients are done using run length coding for the better compression and the bit stream of the coder is stored on any medium or transmitted. At last the bit-stream is formed and transmitted.

3.2 Decoder

The video decoder required at receiver is as shown in Fig. 6:

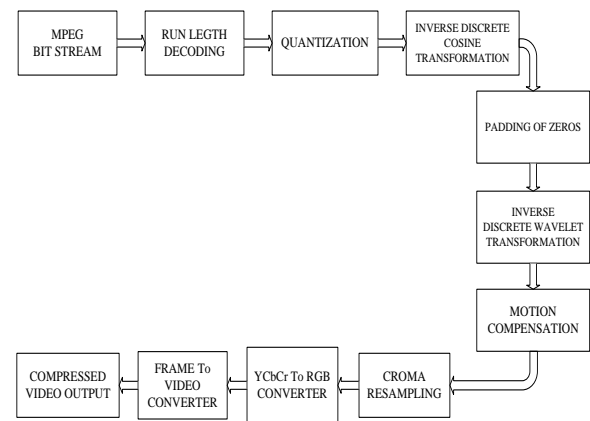


Fig. 7: Modified MPEG-2 video decoder

At the receiver end exactly inverse steps are performed, as the receiver receives the MPEG bit stream the data bits are run length decoded and the quantized coefficients of DCT are achieved. Then these coefficients are de-quantized using the same quantization matrix and inverse DCT is taken of the coefficients, which gives us approximate component of DWT of motion compensated frame. Inverse DWT is taken after padding the zeros for horizontal, vertical and diagonal components, which leads to the motion compensated frames. Then from these GOP the further steps are similar to regular MPEG as explained in previous section for recovery of the original video.

4. VIDEO QUALITY METRICES

The parameter mainly used for the sake of subjective measure of degradation of video after the compression is:

$$PSNR = 10 * \log \frac{255^2}{MSE} \quad (1)$$

The phrase peak signal-to-noise ratio, often abbreviated PSNR [8], is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale.

The other parameter which is used to check subjective quality of compressed video is SSIM (Structural Similarity), which is described as below equation [15]:

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

The structural similarity (SSIM) index is a method for measuring the similarity between two images. The difference with respect to other techniques mentioned previously such as MSE or PSNR, is that these approaches estimate perceived errors on the other hand SSIM considers image degradation as perceived change in structural information. Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close. These dependencies carry important information about the structure of the objects in the visual scene.

Here video quality metric (VQM) used is DCT-based, which calculates mean distortion and maximum distortion in compressed frame with respect to original one and from that VQM as below equation [14]:

$$VQM = (Mean_Dist + 0.005 * Max_Dist) \quad (3)$$

The blurring beta is comparing the blurring artifact of two frames. Lower the value of blurring beta, more the frame is blurred than other one.

The blocking beta is created to measure subjective blocking effect in video sequence. For example, in contrast areas of the frame blocking is not appreciable, but in smooth areas these edges are conspicuous. The value of blocking beta will be more for the frames with more blocking artifacts.

Also the compression ratio is found out between two videos: the original one and the compressed one i.e. bit-stream coming out of video encoder, which is measured by equation as below:

$$CR = \frac{CompressedVideoDataRate}{UncompressedVideoDataRate} * 100 \quad (4)$$

5. COMPARATIVE RESULTS

Here both algorithms are implemented in MATLAB version 7.10.0.499a. For testing purpose seven different standard videos are taken which are having difference in the motion, color component etc. from each other.

Then for the measuring all parameters one popular test bench named, MSU Video Quality Measurement Tool 3.0 has been used. This tool is measuring all the parameters mentioned in the section-4 between the original video and the recovered video at the receiver side of the decoder.

The values shown in the table are average values for each video.

Table. 1: Measured Parameters for different Videos

Name of Video	Name of Parameter	Normal MPEG-2	Modified MPEG-2
vipbarcode.avi	PSNR	29.52	28.86
	MSE	682.46	650.31
	SSIM	0.88	0.84
	VQM	5.37	5.42
	Blocking Beta	16.24	22.24
	Blurring Beta	21.67	18.62
forman.avi	CR	85.93%	92.57%
	PSNR	26.36	24.73
	MSE	238.69	296.66
	SSIM	0.85	0.77
	VQM	3.77	4.23
	Blocking Beta	8.39	20.63
vipmen.avi	Blurring Beta	25.79	20.63
	CR	88.40%	93.83%
	PSNR	24.77	24.51
	MSE	269.24	281.23
	SSIM	0.86	0.82
	VQM	3.83	4.05
viplane.avi	Blocking Beta	14.41	22.28
	Blurring Beta	16.69	16.12
	CR	92.02%	95.34%
	PSNR	23.94	23.59
	MSE	338.07	341.85
	SSIM	0.83	0.77
viplandeparture.avi	VQM	4.53	4.68
	Blocking Beta	12.54	20.06
	Blurring Beta	15.55	13.68
	CR	90.37%	94.77%
	PSNR	26.56	25.98
	MSE	225.20	230.60
rhinos.avi	SSIM	0.89	0.85
	VQM	3.32	3.44
	Blocking Beta	13.35	15.34
	Blurring Beta	10.90	10.47
	CR	91.15%	94.83%
	PSNR	25.94	25.61
viptraffic.avi	MSE	543.03	535.95
	SSIM	0.84	0.79
	VQM	5.93	5.99
	Blocking Beta	10.20	15.94
	Blurring Beta	16.99	16.18
	CR	88.03%	93.45%
vipbarcode.avi	PSNR	24.71	24.42
	MSE	471.41	474.92
	SSIM	0.79	0.73
	VQM	4.94	5.14
	Blocking Beta	14.33	26.42
	Blurring Beta	20.83	16.41
	CR	90.98%	94.97%

6. CONCLUSION

Here the results show that the MPEG-2 encoder gives the compression ratio of around 90%, whereas the modified version of it gives compression ratio of around 95%. So by modifying the MPEG-2 encoder 5% more compression can be achieved. But this compression is not achieved on the loss of quality of video, which can be seen from the other parameters like SSIM, MSE, VQM, Blocking Beta and Blurring Beta. Values of all other parameters are much nearer to that of MPEG-2 that we can say that the quality of video remains as it is, then also we are getting more compression.

7. REFERENCES

- [1] 'MPEG Digital Video-Coding Standards', IEEE Signal Processing Magazine, September, 1997, pp. 82-100.
- [2] White Paper of Axis Communication, 'H.264 video compression standard', 2008.
- [3] White Paper of Axis Communication, 'An explanation of video compression techniques', 2008.
- [4] White Paper of Array Microsystems Inc., 'Video compression-An Introduction', 1997.
- [5] Ian Gilmour and R. Justin Davila, 'Lossless video compression for Archives: Motion JPEG2K and other options', Media Matters IIC.
- [6] P. N. Tudor, 'MPEG-2 Video Compression', Electronics & Communication Engineering Journal, 1995, Paper No. 14.
- [7] Aroh Barjatya, 'Block matching algorithms for Motion estimation', DIP 6620 Spring 2004 Final Project Paper
- [8] MSU-Video Quality Measurement Tool 3.0, www.compression.ru
- [9] PicTools Jpeg 2000, www.jpg.com/jpeg2000
- [10] MPEG-2 Description, www.mpeg.chiariglione.org/standards/mpeg-2
- [11] John G. Apostolopoulos, 'Video Compression', MIT 6.344, Springer 2004.
- [12] Zhou Wang, Eero P. Simoncelli and Alan C. Bovik, 'Multi-scale structural similarity for image quality assessment', Proceeding of 37th IEEE Asilomar conferece on signals, systems and computers, 2003.
- [13] Chaofeng Li and Alan Conrad Bovik, 'Content-weighted video quality assessment using a three-component image model', Journal of Electronic Imaging, 19(1), 011003, Jan-Mar 2010.
- [14] Feng Xiao, 'DCT-based video quality evaluation', Final Project for EE392J, Winter 2000.
- [15] Zhou Wang, Hamid Rahim Sheikh, Eero P. Simoncelli and Alan C. Bovik, 'Image quality assessment: From error visibility to structural similarity', IEEE Transactions on Image processing, Vol. 13, No. 4, April 2004.