

## Video Quality Analysis for H.264 Based on Human Visual System

Subrahmanyam.Ch<sup>1</sup>, Dr.D.Venkata Rao<sup>2</sup>, Dr.N.Usha Rani<sup>3</sup>

<sup>1</sup>(Research Scholar, Dept of ECE, Vignan University, Guntur, India)

<sup>2</sup>(Principal, Narasaraopeta Institute of Technology, Guntur, India)

<sup>3</sup>(Head of the Department, Dept of ECE, Vignan University, Guntur, India)

**Abstract:** - The quality of video, such as the Mean Squared Error (MSE), Weighted Mean Squared Error (WMSE) or the Peak Signal-to-Noise Ratio (PSNR), not always correspond with visual observations. Structural Similarity based video quality assessment is proposed under the assumption that the Human Visual System (HVS) is highly adapted for extracting structural information from video. While the demand on high resolution quality increases in the media industry, Color loss will make the visual quality different. In this paper, we propose an improved quality assessment method by adding color comparison into structural similarity measurement system for evaluating video quality. Then we separate the task of similarity measurement into four comparisons: luminance, contrast, structure and color. Experimental results show that the predicted quality scores of the proposed method are more effective and consistent with visual quality than the classical methods using two different distortion types of video sets.

**Keywords:** - MSE, WMSE, PSNR, HVS, luminance, contrast

### I. INTRODUCTION

Visual information is very important for hearing-impaired people, because it allows them to communicate personally using the sign language. In our research, we focused on the fact that some parts of the person using the sign language are more important than others. We presented a visual attention model based on detection of low-level features such colour, intensity and texture and combination with the prior knowledge – in our case information about skin in image. Information about the visually relevant parts allows us to design an objective metric for this specific case. We presented an example of an objective metric based on human visual attention and detection of salient object in the observed scene.

There is a huge use of semantic information in video. The ability to detect features is an interesting challenge by itself, but it takes on added importance to the extent it can serve as a reusable, extensible basis for query formation and search. Nowadays, the researchers focus mainly on solving the problems of finding the semantic information in video sequences.

### II. INTRODUCTION TO H.264

H.264/AVC is the state-of-the-art video coding standard developed by Joint Video Team (which comprises of ITU-T VCEG and ISO/IEC MPEG). Primary aim of H.264/AVC project was to develop a standard which provides same video quality at about half the bit rates of previous standards like MPEG-2, H.263 etc. Additional goals were lowering the complexity and increasing the flexibility so that the standard can be applied to wide variety of applications over different networks and systems.

Its first version was released in May 2003 under the name H.264/AVC by ITU-T. Since then many versions has been released which has improved the standard in a variety of ways. These new versions are known as Fidelity range extensions (FRExt). The draft for FRExt was completed in September 2004. Scalable video coding (SVC) is another aspect of H.264/AVC project which has gained a lot of popularity. The aim of SVC is to use a single bit-stream which could serve different networks and different devices. SVC extension was completed in November 2007.

In sync with previous standards, H.264/AVC defines only the syntax of encoded bit-stream and the method of decoding it. But this limitation of standard provides maximum freedom to optimize implementations in a manner appropriate to specific applications. Fig 2.1 shows the scope of H.264/AVC standard.

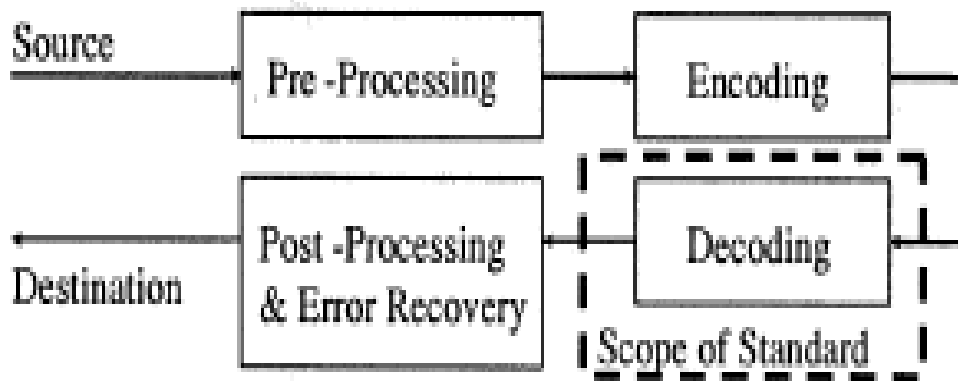


Fig 2.1: Scope of video coding standardization

Since the first video coding standard (H.261), the functional blocks of the basic video coding algorithm remains the same. But the major difference lies in the details of each functional block. As a result each new video coding standard became more comprehensive and complex. Hence the video coding algorithm of H.264/AVC is also similar to previous standards but with an exception of de-blocking filter. Now let's have a look at the codec (enCOder and DECOder pair) of H.264/AVC; starting with the encoder.

**2.1.1 Encoder:**

A video encoder is a device or software which performs compression over the bit-stream; as a result the size of the encoded bit stream greatly reduces in size. Before elaborating more about video encoder, some basic terms of video processing are introduced which will help a lot in explaining various functional blocks of the encoder afterwards.

- Video Stream: A video stream consists of large number of still images, when these images are run at a rate of 25/30 frames per second a video is formed. These still images can be classified as frames or fields depending upon the procedure of scanning.
- Macroblock (MB): Macroblock is special word in video coding used to describe a small region in a frame or field (usually the size of a MB is 16x16 pixels). Therefore, each frame/field is divided into large number of macroblocks.
- Pixel (PICTure ELEment): is the smallest piece of information in a frame/ image.

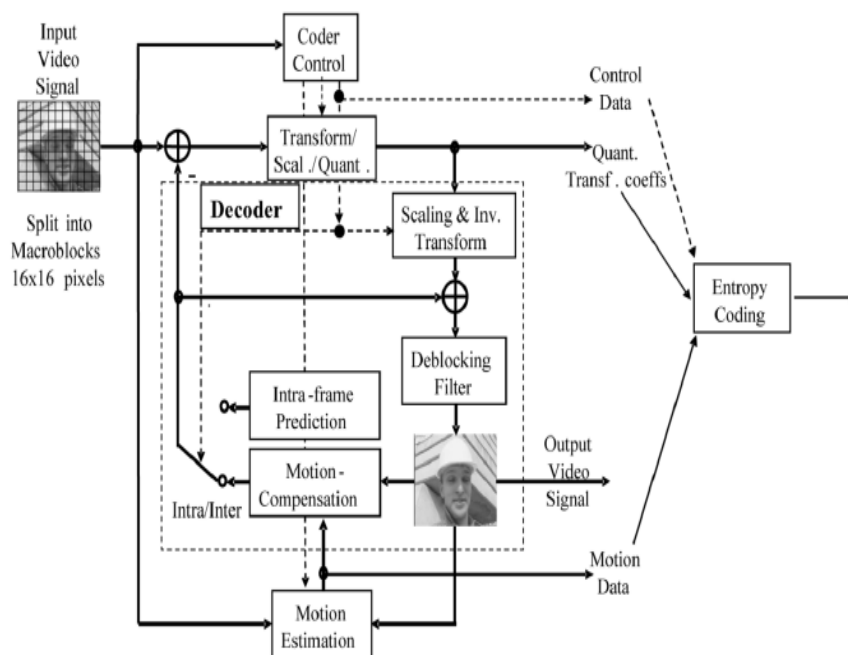


Fig 2.2 Basic encoding structure of H.264/AVC for a Macroblock

Any video coding technique tries to reduce the size of the encoded bit stream by removing the redundancy present in the video sequence. Any video sequence has three types of redundancies temporal, spatial and statistical. To reduce temporal redundancy, similarities between the neighboring video frames are exploited by constructing a prediction of current frame. Normally a prediction is formed using one or two previous or future frames and is improved by compensating differences between different frames. The output of this whole process is a residual frame. These residues are processed by exploiting similarities between the neighboring pixels to remove spatial redundancy. Reduction is achieved by applying a transform (like DCT) and quantizing their results. The transform converts the sample into another domain where they are represented as transform coefficients; these coefficients are quantized to remove insignificant values, thus leaving a small number of coefficient values which provides more compact representation of residual frame. To remove statistical redundancy, an encoder uses various techniques (Huffman coding, Arithmetic coding etc) which assign short binary codes to commonly occurring vectors. Due to the reduction of redundancies the size of encoded bit-stream greatly reduces in size. Therefore, any video encoder consists of these three functional modules.

Fig2.2 shows the H.264/AVC encoder, as is evident from the figure that H.264/AVC Encoder has a decoder loop incorporated in it. The main reason for including the decoder is to provide reference pictures which are very necessary for the prediction process. Therefore the encoder has two data flow paths, forward path and backward or reconstruction path.

- **Forward path:** Here the normal encoding process is executed. An input frame is processed in terms of MB's (16x16). Depending upon the type of MB, prediction technique is selected. There are two types of prediction techniques, intra prediction and inter prediction. In intra prediction, prediction is calculated from the samples of same frame which are previously encoded, decoded and reconstructed, whereas in inter prediction, prediction of current frame is calculated from the Motion compensation prediction of reference pictures which could be either previous or future frames. So when an 'I' MB is processed intra prediction is selected and when a 'P/B' MB is processed inter prediction is selected. Inter prediction techniques relies upon the motion estimation and motion compensation techniques for accurate prediction. These techniques are very instrumental in reducing the temporal redundancy between the frames.

The goal of motion estimation technique is to find a best match for a block of current frame in reference frame. To find the best match, a search is performed in the reference frame. This search is carried out by comparing the block in current frame with the blocks in search area (usually centered on and around the current block position) of reference frame. The block or area which gives the lowest residual energy is considered as the best match region. Thus chosen best match region is called as 'predictor' for the block in current frame. In motion compensation, predictor is subtracted from the current block resulting in residues.

So obtained residues are block transformed, scaled and quantized to obtain quantized coefficients. These quantized coefficients along with control data and motion vector information is entropy coded. Entropy encoded coefficients along with side information necessary for decoding each macro block at the decoder constitutes the encoded bit stream.

- **Reconstruction path:** Here decoding and reconstruction of encoded frames is performed. To reconstruct the encoded frames, quantized coefficients are scaled (De-quantized) and inverse transformed to obtain residues which are added with the prediction to obtain unfiltered frame. This unfiltered frame contains blocking artifacts; blocking artifacts arise mainly due to the usage of block based integer DCT (Discrete Cosine Transform). To smooth out these blocking artifacts, de-blocking filter is applied over the frame. These filtered frames are stored as reference frames.

### **2.1.2 Decoder:**

When the decoder receives the encoded bit stream, it entropy decodes them to produce a set of quantized coefficients which are reordered, scaled and inverse block transformed to obtain residues. Using the header information received in the encoded bit-stream, decoder generates the prediction. Residues are added to the prediction to obtain the unfiltered version of the original frame. Unfiltered frame contains blocking artifacts which deteriorates the video quality.

De-blocking filter is applied over the unfiltered frame to remove the blocking artifacts and enhance the video quality of filtered frame. Filtered frame are stored as reference frames which are utilized in prediction process.

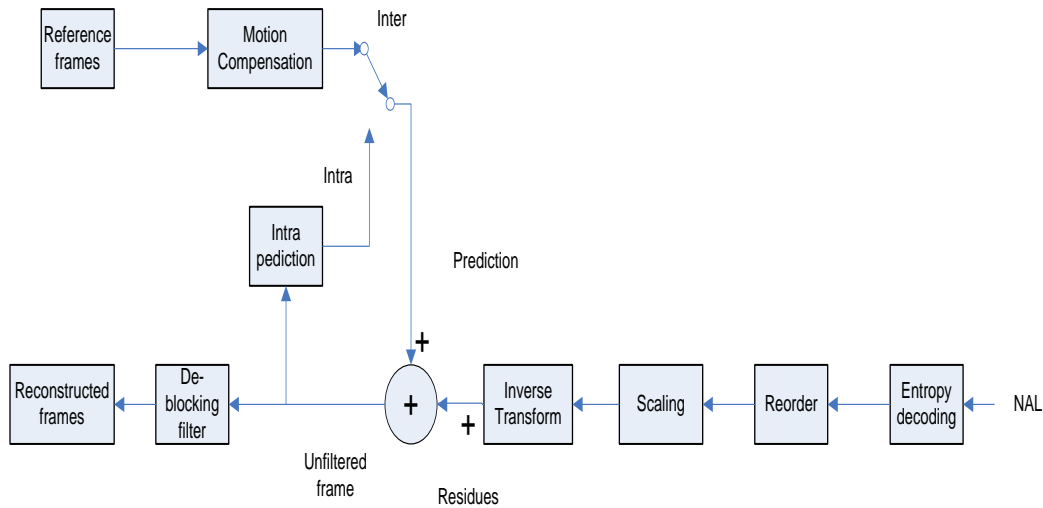


Fig 2.3: H.264 Decoder

### III. PROPOSED REFERNEC MODEL FOR VIDEO QUALITY MONITORING

Three methodologies representing different measurement strategies for the assessment of the quality of video have been defined:

- methodology using the complete video reference (double-ended);
- methodology using reduced reference information (double-ended);
- methodology using no reference signal (single-ended).

The design and the development of a video quality monitor should consider a general structure of the measurement procedure for reduced reference and single-ended methodologies. The reference model is composed of the following four layers:

- *Measurement methodology* defines the class or the strategy relative to the application requirement;
- *Measurement method* is composed of a set of modules, algorithmic and associated ones, implemented to process inputs such as original signals or processed reference data, and provide output results such as processed reference data, level of impairment or final quality notation;
- *Algorithmic module(s)* is the basic block of signal processing functions composing the method. It composes the core of the method from which the final objective qualification is delivered;
- *Associated module(s)* is an additional function that aids the algorithmic module(s) in its operation by addressing such issues as dating, synchronization, presentation of data, etc.

Each one of those methodologies offers a specific approach relative to its operational use. Each methodology is well adapted to a specific field of applications. This is due to its technical implementation constraints:

- relevance of the measured data,
- availability of the reference,
- synchronization of original and impaired signals or data,
- transmission channel for reference data,
- real-time implementation,
- in-service use, etc.

Post-processing algorithms are the most popular methods for improving the quality of the image and video and eliminate the annoying blocking artifact. On the other hand, the post-processing algorithms can achieve deblocking without the original image and video, so the standard need not to be modified.

#### Reduction of Blocking Artifacts in DCT Domain

We introduce a post-processing filtering algorithm in DCT domain. We define the block  $b_{m,n}^{k,l}(u,v)$  and  $B_{m,n}^{k,l}(u,v)$  first.  $b_{m,n}(u,v)$  is the (m,n)-th 8x8 block in the compressed image, and  $B_{m,n}(u,v)$  is the DCT coefficients of  $b_{m,n}(u,v)$ .  $b_{m,n}^{k,l}(u,v)$  is the shifted block with displacement  $k$  pixel in the  $x$  direction and displacement  $l$  pixels in the  $y$  direction with respect to block

$b_{m,n}(u, v)$ , and  $B_{m,n}^{k,l}(u, v)$  is the DCT coefficients of the block  $b_{m,n}^{k,l}(u, v)$ . One example is shown in Fig. 3.1.

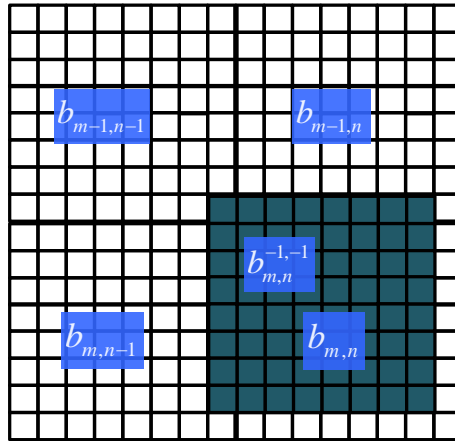


Fig. 3.1 Example of shifted block  $b_{m,n}^{k,l}(u, v)$

In the original image, the neighboring DCT coefficients at the same frequency are very similar and do not vary radically within a small range. Thus, we can apply low pass filter to the DCT coefficients at each frequency to filter the high frequency parts resulting from blocking effect. However, this method may blur the real edges in the original image, so we must have the mechanism to detect activity of the block and apply the filter with corresponding strength.

DCT-domain filtering is applied to revise the block  $B_{m,n}(u, v)$  to obtain the new DCT coefficients

$\hat{B}_{m,n}(i, j)$ .

$$\hat{B}_{m,n}(i, j) = \frac{1}{W} \sum_{k=-h}^h \sum_{l=-h}^h w_{k,l} B_{m,n}^{k,l}(u, v)$$

$$W = \sum_{k=-h}^h \sum_{l=-h}^h w_{k,l}$$

The post-filtering works in different ways for the blocks with different activities.

For blocks with low activity, the blocking artifact is more noticeable, so we apply strong filtering to smooth the high frequency components. The filter coefficients are defined.

$$w_{k,l} = 1, k, l = -2, \dots, 2$$

For blocks with high activity, the blocking artifact is less noticeable, so we apply filtering with less strength to smooth blocking artifact and preserve the real edge. The filter coefficients are defined.

$$w_{k,l} = \begin{cases} 3, & \text{for } (k, l) = (0, 0) \\ 1, & \text{otherwise} \end{cases}$$

#### IV. EXPERIMENTAL RESULTS

In this, we will show some experimental results. The video used in the experiment is “xylophone.mpg”, which consists of original high-resolution YUV format video files. It includes upto 10 frames of re sized video. The details of the video frame values are shown.

Step wise Process:

Encoding I-Frame: 1	Encoding P Frame: 2	Encoding P Frame: 3
Encoding P Frame: 4	Encoding P Frame: 5	Encoding P Frame: 6
Encoding P Frame: 7	Encoding P Frame: 8	Encoding P Frame: 9
Encoding P Frame: 10		



Fig.4.1. Encoding Video process of xylophone.mpg video (10 frames)

Decoding I-Frame: 1  
Decoding P Frame: 4  
Decoding P Frame: 7  
Decoding P Frame: 10

Decoding P Frame: 2  
Decoding P Frame: 5  
Decoding P Frame: 8

Decoding P Frame: 3  
Decoding P Frame: 6  
Decoding P Frame: 9



Fig.4.2. Decoding Video process of xylophone.mpg video (10 frames)



Fig.4.3. PSNR = 18.8015



Fig.4.4. PSNR = 1.2104

## V. CONCLUSION

Emerging new technologies demand tools for efficient indexing, browsing and retrieval of video data, which causes rapid expansion of areas of research where the semantic information is used. New methods that work with semantic information in image, video, and audio are developed frequently these days, which means that our list of methods is not final. Nevertheless, we picked the most used ones and tested them. We brought an overview of the methods for detecting and describing interesting points in video, modeling visual attention and saliency. We have tested the H.264 video formats. We have evaluated the use of visual saliency for compression of video sequences.

**REFERENCES**

- [1] A. C Bovik. Introduction to Image and Video Quality Assessment[OL]. [2009-3-17]. <http://live.ece.utexas.edu/research/quality/intro.htm>.
- [2] C J van den Branden Lambrecht, Ed. Special Issue on Image and Video Quality Metrics [J]. *Signal Process.*, 1998, 70(3): 155-294.
- [3] VQEG. Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment [OL]. (2000-3-15) [2009-3-17]. [http://www.its.bldrdoc.gov/vqeg/projects/frtv\\_phaseI/COM-80E\\_final\\_report.pdf](http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI/COM-80E_final_report.pdf).
- [4] T N Pappas and R J Safranek. Perceptual Criteria for Image Quality Evaluation. In: A. C Bovik, ed. *Handbook of Image & Video Processing* [M]. San Diego: Academic Press, 2000: 669-684.
- [5] Z Wang and A C Bovik. A Universal Image Quality Index [J]. *IEEE Signal Processing Letters*, 2002, 9(3): 81-84.
- [6] G H Chen, C L Yang, L M Po, et al. Edge-based Structural Similarity for Image Quality Assessment [C]. ICASSP, Toulouse, France, 2006: II-933-II936.
- [7] Commission Internationale d'Eclairage. CIE Publication No.15.2, Colorimetry [S]. 2nd Ed. Vienna: CIE Central Bureau, 1986.
- [8] Commission Internationale d'Eclairage. CIE Publication No. TC1. 47, Hue and Lightness Correction to Industrial Colour Difference Evaluation [S]. Vienna: CIE Central Bureau, 2000.
- [9] VQEG. Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II VQEG [OL]. (2003-8-25) [2009-3-17]. [http://www.its.bldrdoc.gov/projects/frtv-\\_phaseII/downloads/VQEGII\\_Final\\_Report.pdf](http://www.its.bldrdoc.gov/projects/frtv-_phaseII/downloads/VQEGII_Final_Report.pdf).