Video Quality Modeling for Quality-driven Design

Cao, Yun
Department of Computer Science and Communication Engineering, Kyushu University

Yasuura, Hiroto
Department of Computer Science and Communication Engineering, Kyushu University

出版情報：The 10th Workshop on System And System Integration of Mixed Technologies(SASIMI 2001), pp.86-92, 2001-10. Workshop on System And System Integration of Mixed Technologies
バージョン：
権利関係：

Video Quality Modeling for Quality-driven Design

Yun Cao Hiroto Yasuura
Department of Computer Science and Communication Engineering
Graduate School of Information Science and Electrical Engineering
Kyushu University
6–1 Kasuga-koen, Kasuga-shi, Fukuoka 816-8580 Japan
{*cao, yasuura*}*@c.csce.kyushu-u.ac.jp*

Abstract— This paper models the video quality, which focuses on the effects of computation precision by combining subjective-objective metric. The motivation behind the research is to bring quality-driven design into effect for video applications. We change the computation-precision of IDCT, the kernel program of MPEG-2 video decoder and get the experimental computation-precision-oriented video quality model. The experimental results show that reducing computation precision while providing certain video quality is a perspective way to reduce cost for embedded system design.

I. Introduction

Modern LSI technology enables implementation of a complex system on a single chip. Progress in LSI technology is extremely fast and it is outstripping the system designers' abilities to make use of the created opportunities. The quality of the system LSI, the power consumption, production cost and time-to-market tends to be more limited by the design methods and tools. Substantial improvement can only be achieved through development and application of a new generation of more suitable design paradigms, methods and tools. In this situation, quality-driven design is presented.

The quality-driven design is a kind of user-application-oriented design, whose process is evolutionary and it basically consists of building the quality models, using them for constructing, selecting and improving the design solutions, which satisfy the users' requirements and are "totally" optimal for applications. The key techniques for quality-driven design are quality modeling, quality measuring, and design decision-making.

In order to bring the quality-driven design into effect, the definition of quality has to be solved. As we known, until now none of the exist quality definition is precise enough to enable the systematic consideration, measurement and comparison of quality, which are necessary for quality-driven design. Most of them focus exclusively on a product being designed and do not account for design. This paper focuses on quality-driven design for video applications, a case study of quality-driven design. As far as we know, this paper first models the video quality, which focuses on the effects of computation precision.

The advent of digital video systems has exposed the limitations of the techniques traditionally used for video quality measurement. For conventional analog video systems there are well-established performance standards. They rely on particular test signals and measurement procedures to determine parameters such as differential gain, differential phase or waveform distortion [4]. While video compression, storage, and transmission system has not invalidated these traditional parameters, it has certainly made their connection with perceived video quality much more tenuous. The designers of compression algorithms have had to resort to subjective viewing tests in order to obtain reliable ratings for the quality of compressed images or video [5]. However, these tests are complex and time-consuming.

Looking for faster alternatives, researchers have turned to simple error measure such as root mean squared error(RMSE) or peak signal-to-noise ratio(PSNR), suggesting that they would be equally valid. However, these simple error measures operate solely on a pixel-by-pixel basis and neglect the important influence of image content and viewing conditions on the actual visibility of artifacts. Therefore, they cannot correlate well with perceived quality, and many experiments confirm this low correlation [6] and [7]. In order to be able to replace subjective rating experiments, the ideal objective quality assessment system should rate video impairments just like a human being, however, it is very difficult by our limited knowledge of the human visual system.

In order to bring the quality-driven design into effect, this paper models the quality of video, which focuses on the effects of computation precision to explore the design spaces from output quality of video applications.

Our paper is structured as follows: the next Section 2 gives motivation and preliminaries of our work. Section 3 models video quality for quality-driven design. Experiments and results are shown in section 4. Finally, section 5 gives conclusion and our future work.

II. Preliminaries

A. Motivation

Because of high complexity, dynamic and multi-aspect character of system design problems, quality cannot be well defined. However, it can and should be modeled. There are currently no accepted means for measuring or quantifying the visual quality of digital video for quality-driven design. To remedy this situation, we model the video quality of MPEG-2 video system by combining subjective-objective metric, which focuses on the effects of computation precision.

For video applications, a huge amount of high quality digital data encoded in standard formats are produced and stored. On the other hand, various kinds of output devices with different characteristics are developed. A digital video system that works fine for family television might be inadequate for video teleconferencing with bigger display, however, it is "redundant" for mobile phone with smaller display. To take the advantage of these characteristics, we propose quality-driven design methodology, which explores flexible decoder systems for video applications to achieve low power and low cost while supply enough performance. In order to bring quality-driven design into effect, we need to model the video quality, which is precise enough to enable the systematic consideration, measurement and comparison.

B. Quality-driven Design for Video Applications

When design embedded systems and SoC (System on a Chip), designers have to consider the trade-off among system performance, cost and power consumption [1]. The data width computed in the system is one of the most important design parameters related with performance, cost and power of the system [2]. The width of datapath and the size of memories strongly depend on the data width. System designers often spend much time to analyze the data width required in the computation of the system [3]. Hardware designers of portable multimedia devices reduce the datapath width (the length of registers and the bit width of operation units). Programmers of embedded systems work hard for adjustment of the width of variables to keep the accuracy of computation. By controlling the datapath width and/or the length of registers, we can reduce the cost and power consumption drastically. Furthermore, we can choose the computation precision really required for each application. In video processing, for instance, the required quality of video, such as resolution and levels of color, strongly depend on the characteristics of output display devices. We can reduce the computation precision in target application program, if the reduction does not induce decrease of output quality. It means that we can design an video system with the minimum hardware and energy consumption by eliminating redundant computation. We call the design methodology, quality-driven design for video applications(QDDV).

The main concepts of QDDV can be formulated as follows:

$$\begin{array}{ll} Given & AnApplicationProgram \\ lookfor & ComputationPrecision \\ subject\ to & MOS(CP) \geq M_{cst},\ PSNR(CP) \geq S_{cst} \\ & E(CP) \leq E_{cst},\ A(CP) \leq A_{cst} \end{array}$$

Where $MOS(CP)$, $PSNR(CP)$, $E(CP)$ (energy consumption), and $A(CP)(area)$ are functions of the computation precision CP, M_{cst}, S_{cst}, E_{cst} and A_{cst} are the constraints on the subjective mean opinion score(MOS), objective peak signal-to-noise ratio(PSNR), energy consumption and area respectively. This is a nonlinear optimization problem. Quality of computation in the system is determined by the trade-off among quality of output, power consumption and cost of the system.

C. Video Quality Factors

It is necessary to understand what "quality" means to a viewer in order to bring the quality-driven design into effect. Viewers' enjoyment when watching a video depends on many factors. Video quality plays a prominent role. Research has shown that video quality depends on viewing distance, display size, resolution, brightness, contrast, sharpness, colorfulness, naturalness and other factors [8] and [9].

$$\alpha = 2\arctan(H/2D) \tag{1}$$

$$f_{max} = L/2\alpha[cpd] \tag{2}$$

It is helpful to relate the definitions of some of quality factors to video quality modeling and human visual system. For instance, in the video community it is very popular to specify viewing distance in terms of display size, i.e. in multiples of screen height. The recent experiment results show that the preferred viewing distance is around 6 or 7 screen heights for smaller displays, it approaches 3 to 4 screen heights with increasing display size [10]. In the context of vision modeling, the size and resolution of the image are more specifications. The size is measured in degrees of visual angle α, and the resolution or maximum spatial frequency f_{max} is measured in cycles per degree of visual angle (cpd). For a given screen height H, viewing distance D and number off scan lines L, these two units are computed in formulation above.

III. Video Quality Modeling

Well-structured models of the required quality are extremely important. They can serve to conceptualize, denote, analyze and communicate the customer and designer's ideas, and also to guide the design process. We employ two measures to model video quality. One is a subjective measure that is derived from the subjective data while the other is an objective measure that is derived

TABLE I
MOS GRADING SCALE

Scale	Impairment
5	Imperceptible
4	perceptible, but not annoying
3	slightly annoying
2	annoying
1	very annoying

TABLE II
MPEG-2 TEST BITSTREAMS

Bitstream name (Abbreviation)	resolution (pixels*lines)	bit rate (Mbit/sec)
flwr015	352*240	1.5
flwr400	704*480	40
mobl015	352*240	1.5
mobl400	704*480	40
susi015	352*240	1.5
susi400	704*480	40
sonyct1	352*224	1.5
sonyct2	704*480	40
tek3	512*512	3.5

from computer-based processing of the sampled video images. The objective measure of video quality is developed using the set of MPEG-2 test bitstreams described in this section. The subjective measure of video quality is calculated by taking the absolute value of the average MOSs for each test bitstream. By combing the subjective MOSs and objective PSNRs, we model the video quality for quality-driven design, which focuses on the effects of the computation precision.

A. Subjective Measure

The visual or subjective assessment of video quality has drawn attention of a number of researchers for many years, principally in relation to evaluation of new transmission or coding schemes, and in the development of advanced television standards. The standardization committees of the ISO, and in particular the CCIR, have published recommendations on the assessment of picture quality in television.

Subjective experiments are necessary in order to evaluate models of human vision. Similarly, subjective ratings form the benchmark for visual quality metrics. Formal subjective testing is described in ITU-R Recommendation 500, which suggests standard viewing conditions, criteria for observer selection, assessment procedures, and analysis methods.

In our work, we follow closely the CCIR 500 recommendations with respect to subjective scales and experimental conditions. We make use of the numerical scores associated with the impairment descriptors, or categories of table 1, which show the 5-point (MOS) impairment scale for the computation of average MOS scores.

Because the subjective perception of noise and the behavior of MPEG-2 systems are influenced by scene attributes such as spatial detail, amount and complexity of motion, brightness, and contrast, test scenes that spanned a range of these attributes are selected. MPEG-2 video bitstreams of Table 2 are used as test scenes.

The specific conditions used are as follows:

- 1:The pictures used are four kinds of resolutions and are viewed at four times the picture height (4H).
- 2:Non-expert viewers participate in the experiments, and individual ratings are recoded.
- 3:Nine clips of MPEG-2 test bitstreams are used.

In the experiments, the observers are asked to assign a score A(i,k) to each test bitstream, where A(i,k) is the score given by the i_th observer to test bitstream k. The scores are averaged to obtain the MOS value for specific image.

$$MOS(k) = \frac{1}{n}\sum_{i=1}^{n} A(i,k) \tag{3}$$

where n denotes the number of observers.

We changed the computation precision of IDCT, a kernel program in MPEG-2 video decoder and got the results of figure 2, which show the relationship among the internal variable bit length in IDCT, MOSs and clips for four kinds of resolutions. Here, each point in the graph represents the average quality of a MPEG-2 video decoder system, which was obtained by averaging the subjective MOSs. Figure 4 shows the relationship among the internal variable bit length in IDCT, PSNRs and clips for four kinds of resolutions. In figure 3 and figure 5 over all scenes(nine bitstreams) were injected into the MPEG-2 video system.

Subjective assessment tests are widely used to evaluate the picture quality of coded images, but careful subjective assessments of quality are experimentally difficult and lengthy, and the results obtained may vary depending on the test conditions. Further, subjective assessments provide no constructive methods for performance improvement, and are difficult to use as part of the design process. However, the most fundamental quality measures for digital video are the subjective responses of human viewers to delivered images and subjective tests remain the only viable reference point for validating objective measures.

B. Objective Measure

Objective measures of picture quality can make the comparison of coded images, and also have the possibility of successive adjustments to improve or optimize the picture quality for a desired quality of service. The objective assessment of performance both with respect to bit rate and image quality would also lead to a more systematic design of video systems. An objective model that produces overall quality estimates would have to account for application-specific effects. The influence on modeling accuracy is the changing expectations of people over time. For these reasons, objective video quality modeling is valid only if the application and viewer population are well defined.

PSNR is often used to specify the signal-to-noise ratio of a video signal. This method has the advantage of removing the signal power, which varies from scene to scene from the signal-to-noise-ratio (SNR) calculation so that a given SNR is indicative of some fixed amount of noise power. We calculate PSNR according to the following formulation:

$$PSNR = 10 \times \log_{10}[\frac{1}{E} \times 255^2][dB] \quad (4)$$

where, $PSNR$: Ratio of peak signal to noise
E : Mean-square error

It is important that an objective scale mirror the perceived video quality. For instance, simple distortion scales, such as the signal to noise ratio (PSNR), or even the weighted mean square error (WMSE) are good distortion indicators for random errors, but not for structured or correlated errors. There have been many studies of the construction of objective scales, which represent properties of the human observer [11]. Compression algorithms used in digital video systems produce artifacts whose visibility strongly depends on the actual video content. Simple error measures such as RMSE or PSNR, albeit popular, ignore this important fact and are only a mediocre predictor of perceived quality. Many applications require more reliable assessment methods.

C. Modeling Video Quality for QDDV

The subjective quality perceived by the users that determines whether an application is adopted. The ultimate benchmark would be for objective measures to replace subjective experiments altogether. We model video quality by conducting simultaneous subjective and objective tests. This model building process is necessary for determining the overall accuracy of the objective parameters and for identifying the portion of the subjective responses explained by the objective parameters.

The video quality modeling for quality-driven design is conducted by the following phases:

- Phase 1: Extracting statistics data from subjective measures MOSs for different computation precisions.
- Phase 2: Extracting statistics ata from objective measures PSNR for different computation precisions.
- Phase 3: Calculating a composite video quality model by combining subjective measure and objective measure considering computation precision.

$$Cmdl_{QDDV} = Comp(MOS, PSNR) \quad (5)$$

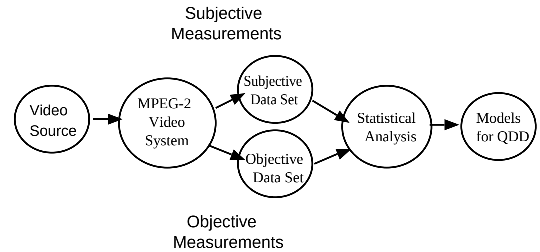

Fig. 1. Process of building video quality model for QDDV

- Phase 4: Modeling the video quality for QDDV.

$$MDL_{QDDV} = Rank(Cmdl_{QDDV}) \quad (6)$$

Figure 1 illustrates the process that is being used to model the video quality for QDDV. Consistent with the need for quality-driven design, the video source material was selected to be representative of the actual end-user's applications. This work used MPEG-2 test bitstreams. The source material is passed through MPEG-2 video systems. The resulting impaired destination material is evaluated in subjective tests. Objective measures, extracted from the digitized video signal, are compared to the subjective test responses using statistical analysis techniques. Only objective measures that accurately predict the subjective responses over a suitable range of test conditions are considered. Otherwise, the subjective results are used.

Objective model results are compared with subjective data to determine the model's performance. To be of value, perception-based objective quality measures must be well correlated with subjective viewer responses. Tests of correlation between objective measures and subjective responses should been conducted in a number of video experiments. The perception-based features are measured using the same source and destination video as the subjective tests. Objective quality parameters, derived from the perception-based features, are than used to predict the subjective viewer responses.

Statistical analysis techniques such as coefficients of correlation and analysis of variance (ANOVA) provide the fundamental tools for determining how much and what potion of the total subjective variance can be explained by objective parameters.

There are a number of attributes that characterize a visual quality metric in terms of its estimation performance with respect to the subjective ratings. These attributes are accuracy, monotonicity and consistency. Accuracy is the ability of a metric to predict subjective ratings with minimum average error and can be determined by means of the Pearson linear correlation coefficient; for a set of N data pairs (x_i,y_i),it is defined as follows:

$$r_p = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2}\sqrt{\sum(y_i - \overline{y})^2}} \quad (7)$$

where $\overline{x}$ and $\overline{y}$ are the means of the respective data sets.

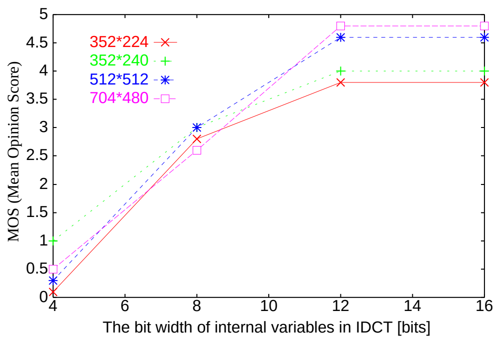

Fig. 2. MOSs of 4 kinds of resolutions for different internal variable bit length in IDCT

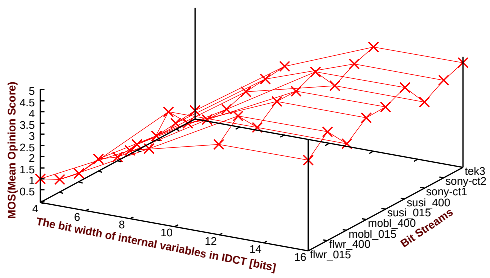

Fig. 3. MOSs of bitstreams for different internal variable bit length in IDCT

Monotonicity is another important attribute. Ideally, changes of a metric's rating between different sequences should always have the same sign as the changes of the corresponding subjective ratings. The degree of monotonicity can be qualitified by the Spearman rank-order correlation coefficient, which is defined as follows:

$$r_S = \frac{\sum(\lambda_i - \overline{\lambda})(\gamma_i - \overline{\gamma})}{\sqrt{\sum(\lambda_i - \overline{\lambda})^2}\sqrt{\sum(\gamma_i - \overline{\gamma})^2}} \tag{8}$$

$$= 1 - \frac{6(\lambda - \gamma)^2}{N(N^2 - 1)} \tag{9}$$

where λ is the rank of x_i and γ_i is the rank off y_i, and $\overline{\lambda}$ and $\overline{\gamma}$ are the respective midranks.

IV. Experiments and Results

This section presents experiments and results. Measurements of MPEG-2 video decode system by changing the computation precisions of IDCT, the kernel program of MPEG-2 video decoder have been conducted individually for video quality modeling. This study investigated the quality effects of computation precision on an MPEG-2 video decoder system. We use the modeling method presented in the previous section. The subjective experimental setup for the measurements based on ITU Rec.500 as described previously. The input video material included some test scenes from the original Rec. 601 test. Table 2 gives a description of the nine scenes that are used for the experiments. The input test scenes are chosen for their variety, although they do not necessarily represent the full range of video of interest.

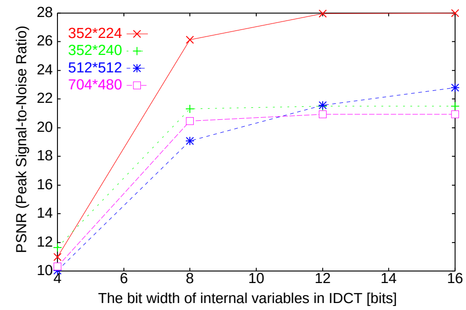

Fig. 4. PSNRs of 4 kinds of resolutions for different internal variable bit length in IDCT

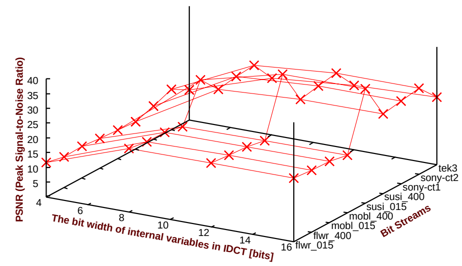

Fig. 5. PSNRs of bitstreams for different internal variable bit length in IDCT

Mpeg2decode C source program from the MPEG Software Simulation Group was used. It is a player for MPEG-1 and MPEG-2 video bitstreams. Mpeg2decode is an implementation of an ISO/IEC 13818-2 decoder, whose emphasis is on correct implementation of the MPEG standard and comprehensive code structure. The MPEG-2 core consists of several function blocks such as IDCT, a couple of motion estimation blocks, a motion compensation block, variable length encoding, decoding blocks and so on. IDCT is the kernel part of Mpeg2decode for computation, it consists of three functions, which are *idct*() of two dimension IDCT with 11 lines, *idctrow*() of row IDCT with 54 lines and *idctcol*() of column IDCT with 54 lines.

The primary purpose of the subjective experiment is to collect subjective viewer response data that can be used to construct an objective model of video quality for MPEG-2 video systems. There are four experimental variables that contributed to the variability in the subjective data:(1)test scene, (2)computation precision, (3)coding bit-rate, and (4)viewer. In order to examining the effect of computation precision on perceived quality, we change the computation precision of IDCT, a very import part in MPEG-2 video decoder system. Three kinds of bit-rate are used. Viewers are given the task of rating the difference in quality between the input and output video. The subjective measure of video quality is calculated by taking the absolute value of the averaged MOSs for each test scene.

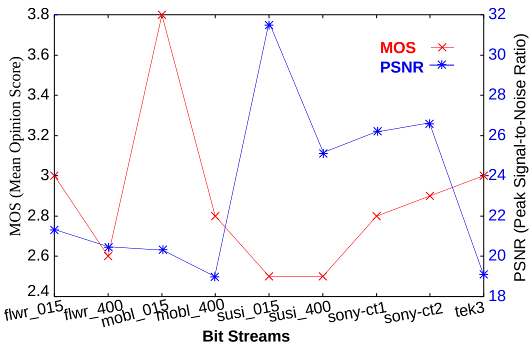

Fig. 6. MOSs and PSNRs of bitstreams for internal variable at 8bit length in IDCT

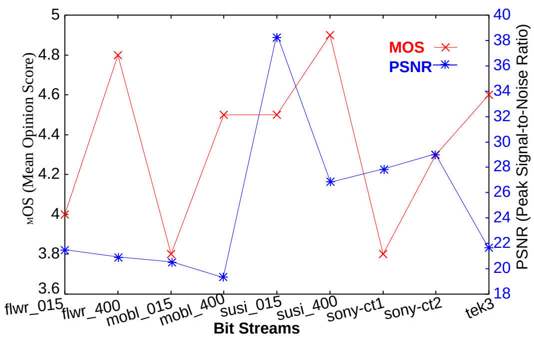

Fig. 7. MOSs and PSNRs of bitstreams for internal variable at 12bit length in IDCT

The objective measure of video quality, which is developed using a set of ANSI standardized test scenes is evaluated using MPEG-2 test scenes described in Table 2. We changed the computation precision of IDCT, computed PSNR value for the nine test scenes and got the figure 5. Figure 6 presents the comparison results of $MOSs$ and $PSNRs$ for different bitstreams at 12bit precision of IDCT. Figure 7 presents the comparison results for 8bit computation precision of IDCT. Here, higher $MOSs$ and higher $PSNRs$ indicate less impairment.

These experimental results indicate that the subjective and the objective measurement errors are random with respect to scene content, and thus we can realize an improvement in measurement accuracy by averaging the subjective scores (MOSs) across viewers and scenes, and by averaging the objective scores (PSNRs) across scenes. From Figure 6 and Figure 7, the representative experimental results for our computation-precision-oriented video quality models of MPEG-2 video decoder system, we can see that the computation precision of 12bit for IDCT may be enough since we can not see any damage in pictures like Figure 9 compared with Figure 8 shown the two frames from original bitstream, and Figure 10 shows the two frames for internal variable at 8bit length in IDCT.

V. Conclusions and Future Work

This paper has modeled the video quality for quality-driven design through combining subjective model, MOS and objective model, $PSNR$. We studied the effects of computation precision on video quality by changing computation precision. Experimental results show that reducing computation precision while providing certain video quality to design embedded system is possible, and we believe that this research is perspective because it can reduce a lot redundancies, which results in drastically reduction of cost including energy consumption and areas of hardware. The theory model for quality design $Cmdl_{QDDV}$ is our future work. We also plan to study analysis algorithm of computation precision and build quality-driven design methodology in the future.

Acknowledgements

This research was partly supported by the Grant-in Aid for Scientific Resarch (B) (2) 12558029 and VCDS project of STARC.

References

[1] Clark N. Taylor, Sujit Dey, and Debashis Panigrahi, "Energy/Latency/Image Quality Tradeoffs in Enabling Mobile Multimedia Communication", *Proc. of Software Radio: Technologies and Services*, Enrico Del Re, Springer Verlag Ltd., pp. 55-66, January 2001.

[2] Marc-Andre Cantin and Yvon Savaria,"An Automatic Word Length Determination Method", *Proc. of The IEEE International Symposium on Circuit and Systems*, V53-V56, May. 2001.

[3] H. Yamashita, H. Yasuura, F. N. Eko, and Yun Cao, "Variable Size Analysis and Validation of Computation Quality", *Proc. of Workshop on High-Level Design Validation and Test*, HLDVT00, pp.95–100, Nov. 2000.

[4] W. Y. Zou, "Performance evaluation: From NTSC to digitally compressed video", *SMPTE J.* 103(12), 795-800, 1994.

[5] ITU-R Recommendation BT.500-7, "Methodology fro the subjective assessment of the quality of television pictures", *ITU*, Geneva, Switzerland, 1995.

[6] J. Lubin, D. Fibush, "Sarnoff JND vison model", *T1A1.5 Working Group* Document 97-612, T1 Standards Committee, 1997.

[7] S. Pefferkorn, J. L. Blin, "Perceptual quality metric of color quantization errors on still images", in *Proc. SPIE*, Vol. 3299, 210-220, San Jose, CA, 1998.

[8] S. A. Klein, "Image quality and image compression: A psychophysicist's viewpoint",in *Digital Images and Human Vision*, ed. A. B. Watson, 73-88, MIT Press, 1993.

[9] J. B. V. Kayargadde, "Image quality prediction in a multidimensional perceptual space", in Proc. *ICIP, vol. 1, 877-880*, Lausanne, Switzerland, 1996 pp.173-178, Aug.1998.

[10] M. Ardito et al., "Preferred viewing distance and display parameters", *MOSAIC Handbook*, 165-181, 1996.

[11] CCIR,"Rec. 500-2, Method for the subjective assessment of the quality of television pictures", In *Recommendations and reports of the CCIR*, Geneva, 1982.

[12] F. X. J. Lukas and Z. L. Budrikis, "Picture quality prediction based on a visual model", *IEEE Transactions on Communications*, vol. COM-30, no. 8, pp.1679-1692, July 1982.

Fig. 8. Two frames from the original bitstream

Fig. 9. Two frames for internal variable at 12 bit length in IDCT

Fig. 10. Two frames for internal variable at 8bit length in IDCT