

# Human Visual System Based Adaptive Inter Quantization

*Jin Li<sup>1</sup>, Jari Koivusaari<sup>1</sup>, Jarma Takala<sup>1</sup>, Moncef Gabbouj<sup>1</sup> and Hexin Chen<sup>2</sup>*

Department of Information Technology, Tampere University of Technology  
Tampere, FI-33720, Finland<sup>1</sup>

School of Communication Engineering, Jilin University  
Changchun, 130022, China<sup>2</sup>

## ABSTRACT

Block effect is one of the most annoying artifacts in digital video processing and is especially visible in low-bitrate applications, such as mobile video. To solve this problem, we propose an adaptive quantization method for inter frames that can reduce visible block effect in DCT-based video coding. In the proposed method, a set of quantization matrices are constructed before processing the video data. Matrices are constructed by exploiting the temporal frequency limitations of human visual system. The method is adaptive to motion information and is able to select an appropriate quantization matrix for each inter-coded block. Based on the experimental results, the proposed scheme can achieve better subjective video quality compared to conventional flat quantization especially at low-bitrate application. Moreover, it does not introduce extra computational cost in software implementation. This method does not change standard bitstream syntax, so it can be directly applied to many DCT-based video codec. Potential application could be for mobile phone, palm computer and other digital devices with low-bitrate requirement.

**Keyword:** video coding, quantization, human visual system, discrete cosine transform (DCT)

## 1. INTRODUCTION

The human eye is fairly good at seeing small differences in brightness over a relatively large area, but not so good at distinguishing the exact strength of a high frequency brightness variation. This fact allows one to get away with a greatly reduced amount of information in the high frequency components. This is done by simply dividing each component in the frequency domain by a constant for that component, and then rounding to the nearest integer, which is usually named quantization in data compression. As the main lossy operation in image/video coding, quantization truncates many of the higher frequency components to zeros and most of the rest become small positive or negative numbers. Since the quantization can remove the majority of the transform coefficients with low energy, high compression performance can be achieved.

Nowadays, most video coding standards such as H.263 and MPEG-1,2,4 utilize two quantization strategies [1]. One, non-uniform quantization is used for intra-coded frames, i.e., I-frames. The non-uniform quantization is always implemented with continuously changing quantization step size and higher frequency coefficient tends to be quantized by larger quantization value. In this way, high compression efficiency is obtained, while the distortion incurred due to the quantization is neglectable. Two, uniform flat quantization that attenuates all frequencies by the same amount is applied for inter-coded frames, i.e., P- and B- frames. Recently, new quantization methods taking advantage of human visual characteristics to obtain visually better reconstruction of video data have been proposed. In [2], a quantization scheme based on human visual system (HVS) is presented. The scheme exploits the fact that HVS is more sensitive to artifacts around the edge areas in frames than those around texture areas. Thresholds are used to classify the local regions based on the importance, texture, contrast and so on. However, the decision of threshold is empirical and the computational complexity greatly increases. Malo *et. al.*[3] proposed a multigrid motion compensation video coding scheme based on HVS. However, it consumes a lot of memory and is not suitable for digital applications with restrict battery lifetime such as mobile video. In a related work, a quantization step selection scheme for intra-frames is presented by Zhou *et. al.*[4]. Wang *et. al.*[5] tried to merge discrete cosine transform (DCT) and quantization into a single operation, but only uniform quantization was applied and thus it can be directly applied to intra-frames in most video coding standards. In [6], the author proposed a new inter quantization scheme by exploiting the

HVS property to improve the compression efficiency. Similarly, the improvement of compression performance is at cost of the highly increasing computational complexity.

In what follows we propose an adaptive inter quantization algorithm focusing on improving subjective video quality. The algorithm takes advantage of a set of quantization matrices that can be initialized beforehand. During the processing, each inter-coded block is adaptively quantized based on the HVS sensitivity to different spatiotemporal frequencies under a certain velocity. The proposed model can efficiently reduce visual artifacts and improve the subjective video quality with almost the same objective video quality. Moreover, it does not increase the computational complexity and only a bit extra memory is required.

Researchers usually prefer objective quality to subjective quality for video quality measurement. The measurement of video objective quality always needs “original” video material. However, the ultimate purpose of video processing is for human viewing. In most cases, mass audiences have no chance to evaluate video objective quality, since “original” video material is not available. They can only evaluate video subjective quality. In these cases, subjective quality is as important as objective quality. So if we achieve higher subjective quality with almost the same objective quality, both researchers and mass audiences will be satisfied.

The rest of this paper is organized as follows. In Section 2, the characteristics of DCT and HVS are briefly reviewed to reveal their relationship to quantization. In Section 3, the algorithm and implementation are proposed. The experimental results are presented in Section 4. Finally, Section 5 concludes the paper.

## 2. SPATIAL CHARACTERISTICS OF DCT AND HVS

DCT [7] maps a set of data into its spatial frequency components. The resulting array of number contains the same number of values as the original array. The first element in the result array is a simple average of all the samples in the input array and is referred to as DC coefficient. The remaining elements in the result array each indicate the amplitude of a specific frequency component of the input array, and are known as AC coefficients. The frequency content of the sample set at each frequency is calculated by taking a weighted average of the entire set. These weighted coefficients are like a cosine wave, whose frequency is proportional to the result array index as shown in Fig. 1. For an 8×8 DCT block, spatial frequencies are distributed from the lowest to the highest frequency and from the upper-left to the lower-right corner of the block as shown in Fig.2. As human visual system is not sensitive to distinguish the exact strength of a high frequency brightness variation, quantization is utilized to remove high frequency components and obtain a high compression ratio without notable visual degradation.

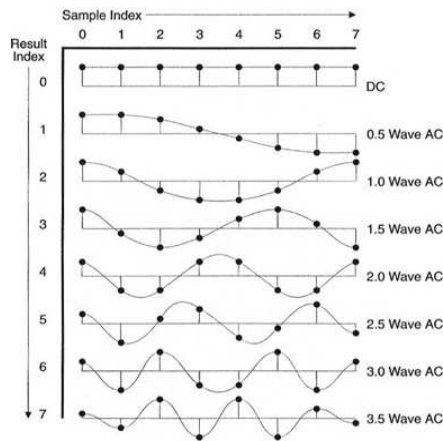


Fig.1 The weighted circle of spatial frequency coefficients in 1-D DCT [7]

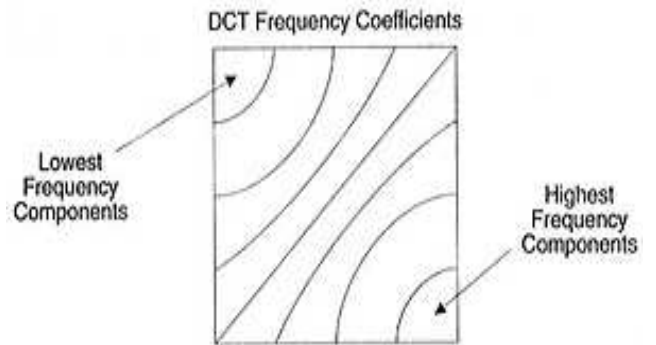


Fig.2. Spatial frequency distribution in DCT domain [7]

In practice, HVS has different sensitivity to different spatial frequencies in DCT domain. D.H. Kelly [8] has proposed an equation to characterize the limitations of HVS to spatiotemporal response

$$G(\alpha, \nu) = \left[ 6.1 + 7.3 |\log(\nu/3)|^3 \right] \times \nu \alpha^2 \exp \left[ -2\alpha(\nu+2)/45.9 \right] \quad (1)$$

where  $v$  is measured in degrees per second and  $\alpha$  is spatial frequency in circles per degree.

Fig.3 is a perceptive view of the threshold surface for the spatiotemporal response characteristics of HVS, where the spatiotemporal frequency response of HVS behaves like a low-pass filter. This low-pass characteristic of HVS means that beyond a certain point HVS is not able to discriminate any spatial variation in a given block. Moreover, the sensitivity of HVS is greatest in the intermediate region. At higher velocities, the sensitivity decreases with increasing velocity, until even a high-contrast target is fused. At low velocities, the sensitivity decreases with decreasing velocity. Eventually, the target begins to fade out, and may disappear entirely at zero velocity, if the stabilization is precise enough.

In conclusion, different spatial frequencies are regularly distributed in DCT domain and the HVS temporal sensitivity differs in terms of spatial frequency and velocity. Therefore a weighted quantization that approximates temporal sensitivity of HVS can achieve better subjective video quality than the flat quantization.

### 3. PROPOSED QUANTIZATION ALGORITHM

#### 3.1 Analysis of proposed quantization algorithm

The velocity of each macroblock is represented by horizontal and vertical velocities,  $v_H$  and  $v_T$ , respectively. The spatial frequencies are characterized by  $\alpha_i$  and  $\alpha_j$ . There are eight frequencies in each direction in an  $8 \times 8$  DCT block, thus  $i$  and  $j$  ranged from 0 to 7 as an integer.

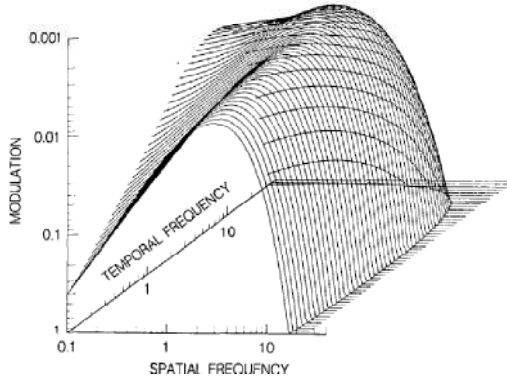


Fig.3. Perspective view of the spatiotemporal threshold surfaces [8]. Each individual curve represents the spatial frequency response at a fixed temporal frequency.

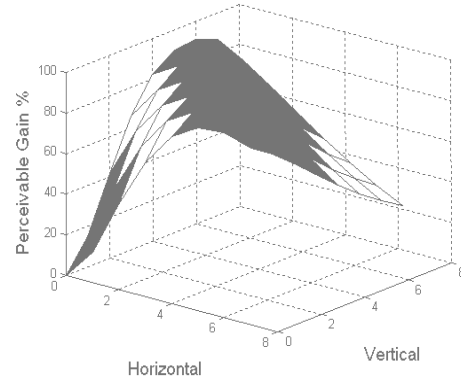


Fig.4. Temporal frequency characteristics of HVS in macroblock. X- and Y-axis represent Fig.3. Temporal frequency character DCT coefficients in horizontal and vertical and Z-axis represents the perceivable gain by HVS.

Then, each temporal frequency response of HVS for two dimensional frames can be calculated from (1)

$$G(\alpha_{ij}, v) = \left[ 6.1 + 7.3 |\log(v/3)|^3 \right] \times v \alpha_{ij}^2 \exp \left[ - 2\alpha_{ij} (v + 2) / 45.9 \right] \quad (2)$$

by first defining

$$v = (v_H^2 + v_T^2)^{1/2} \quad \text{and} \quad \alpha_{ij} = \alpha_i + \alpha_j \quad (3)$$

For an  $8 \times 8$  DCT block, we obtain totally 64 components that are defined as  $G(\alpha_{ij}, v)$ .

In fact, the temporal frequency of the observed object can be regarded as a function of velocity and spatial frequency of the object, which has been confirmed by Dong [9].

In order to calculate the spatial frequency and the velocity for each inter-coded macroblock, we introduce the parameters in accordance to the work in [6].

$$v_H = \frac{m_H}{w} \times f \quad \text{and} \quad v_T = \frac{m_T}{h} \times f \quad (4)$$

$$\alpha_i = \frac{m_H}{m} \times \frac{w}{m} \times c_i \quad \text{and} \quad \alpha_j = \frac{m_T}{m} \times \frac{h}{m} \times c_j \quad (5)$$

where  $m_H$  and  $m_T$  are the absolute values of motion vector  $MV(m_H, m_T)$  in horizontal and vertical directions,  $w$  and  $h$  are the width and the height of the frame respectively,  $f$  represents the frame rate,  $m$  is the size of macroblock. Since in an  $8 \times 8$  DCT block, the circles ranged from 0 to 3.5 with an even interval of 0.5 both in horizontal and vertical directions shown as Fig. 2 in [7], we define

$$c_i = 0.5i \quad \text{and} \quad c_j = 0.5j \quad i, j = 0, 1, \dots, 7 \quad (6)$$

Fig.4 depicts the model of the temporal frequency characteristics of HVS for  $MV(4,4)$  when QCIF size video sequence with 30fps (frames per second) was processed. For this model, size of macroblock was  $16 \times 16$  and the search area used in motion estimation was  $32 \times 32$ .

From Fig.3, we can see how important each spatial frequency is for HVS in a macroblock. Further, we can approximate the spatial limitations of HVS to construct a weighted quantization matrix. The constructed matrix can remove more information

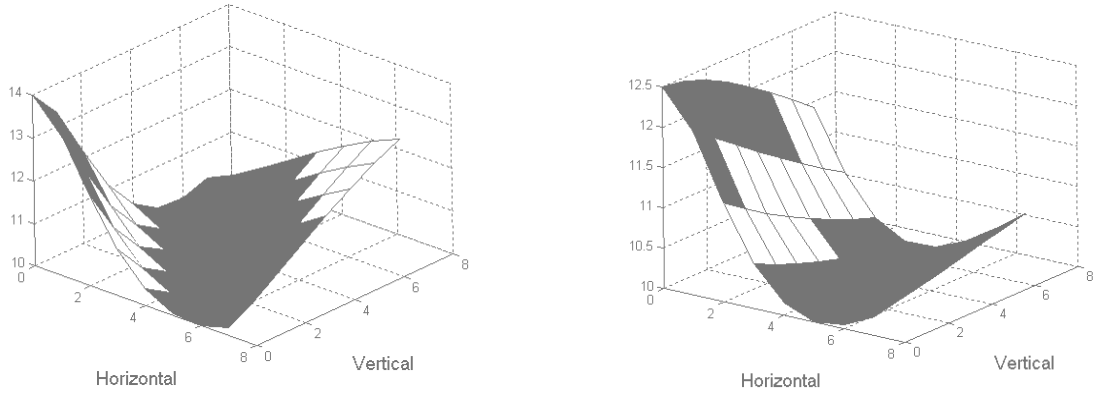


Fig.5. Quantization with different MV. Left:  $MV(4,4)$ ; Right:  $MV(4,1)$ .

than the flat quantization matrix because the HVS is not sensitive to this information. We propose an HVS based quantization model

$$Q_{HVS}(i, j) = (m_H + m_T) / p \times \left(1 - \frac{G(\alpha_{ij}, v)}{G_{\max}}\right) \quad (7)$$

where  $Q_{HVS}(i, j)$  represents a component of the HVS based quantization matrix,  $Q_{HVS}$ ,  $p$  is the adjustment parameter and  $G_{\max}$  denotes the maximum value among  $G(\alpha_{ij}, v)$ .

The proposed model in (7) indicates the relationship between the characteristics of HVS and quantization. Each transform coefficient in a block is quantized based on its sensitivity to HVS. Furthermore, the motion information, i.e.  $MV$ , is used to adjust the quantization parameter for each macroblock. Adaptation to motion information enables the proposed method to efficiently reduce the unnecessary information without affecting to subjective quality.

The proposed quantization method can be obtained by adding the HVS based quantization matrix into the flat quantization matrix

$$Q = Q_{FLAT} + Q_{HVS} \quad (8)$$

Fig.5 shows the proposed quantization matrices obtained from (8) with  $p = 2$  and each component of matrix  $Q_{FLAT}$  equals to 10.

### 3.2 Implementation of proposed algorithm

If the method is directly applied to each inter-coded macroblock, a mass of computations will be introduced. One can see that (3) has a symmetric property, which means that if  $m_H$  and  $m_T$  are exchanged, the operation will only transpose the quantization table. Moreover, the signs of  $m_H$  and  $m_T$  do not affect to resulting quantization table.

If the frame rate, the size of macroblock and the search area are fixed, only a few of the proposed quantization matrices need to be constructed. For the aforementioned video sequence with a  $32 \times 32$  search area, only 35 HVS based quantization matrices are needed. So, the matrices can be initialized beforehand and the best matrix can be selected during the quantization process. In this way, a lot of computations can be saved with slight increase in memory usage.

The method is applied for P- and B-frames. Since only forward MV is used for inter-coded P frames, the proposed model can be applied directly. While for B frames, a distance-based weighting strategy is used. The HVS based quantization is the weighted sum of the two proposed quantization models derived from forward MV and backward MV.

## 4. EXPERIMENTAL RESULTS

### 4.1 Subjective Evaluation

The proposed quantization method was applied to H.263 codec. The method is aimed to improve subjective quality without reducing the objective quality. Therefore, the obtained results are focused on subjective evaluations. Good guidelines for subjective evaluation can be found from Recommendation ITU-R BT.500-11 [10].

During each comparison, two video sequences were displayed synchronously on the screen. One was encoded and decoded by the modified codec. The other was processed by the standard codec. However, observers did not know which sequence was processed by the modified codec. They simply gave their evaluations based on their own perceptual experience. After each comparison, the on-screen locations of the original codec and the modified codec was exchanged randomly to avoid possible influences on the perceptions caused by spatial location differences. Finally, all the evaluation results were combined into our final result. This evaluation was carried out by 16 laypeople whose majors had nothing to do with video technology. Should subjective evaluation be taken by laypeople or professional is a controversial problem. But in recent years, more researchers prefer laypeople. As we know, laypeople constitute the main part of the end users in most cases.

Foreman test video sequence was mainly used because it contains different type of local motion activity. The standard codec and the modified codec were used to produce two decoded video sequences under the same bit rate by adjusting quantization parameter (if not exactly the same, we choose a little lower bitrate for the proposed model). In the experiments, the frame rate was 30 fps, macroblock size and the search area were  $16 \times 16$  and  $32 \times 32$ ,  $p$  in (7) was 2, and no frame was skipped during the processing. Table I list the subjective evaluation results.

Based on opinions of 16 observers, the proposed scheme could achieve better subjective video quality if compared to the standard coding schemes. Table I show that most observers think that the visual quality of the proposed method was better especially at low-bitrate application.

TABLE I SUBJECTIVE EVALUATION WITH H.263

Bitrate (kb/s)	H.263 is better	Proposed is better	Almost the same
57.05	2	12	2
68.70	2	11	3
81.00	3	8	5
114.56	5	8	3
176.90	4	6	6
266.15	3	3	10

If compared to standard codec, the proposed mode tends to allocate more bits to important spatial frequencies and less bits to the frequencies that are not so sensitive to HVS, without changing the overall bitrate. Normally, bit allocation is more influential on visual quality at low-bitrate than at high-bitrate. For instance, the proposed scheme can decrease visible block artifacts by assigning more bits to the most sensitive frequencies in low-bitrate applications. Fig.5 shows the differences in reconstructed frames between the proposed scheme and the standard codec. In Fig.6 since the areas within ellipse have low

motion and are more sensitive to HVS, the modified codec was using finer quantization to these areas. Therefore block artifacts are decreased, as shown in (c) and (d). The rectangle area contains fast hand movement, thus spatial frequencies in these blocks were attenuated more, as shown in (e) and (f). However, observers can't perceive any degradation if the video is played at normal frame rate due to the HVS limitations. Similarly, Fig.7 illustrates the comparison between the proposed quantization algorithm and the baseline H.263 standard in frame 247. Since it only includes low motion activity and is more sensitive to human visual system, frame 247 tends to be located more bits than in the standard codec.

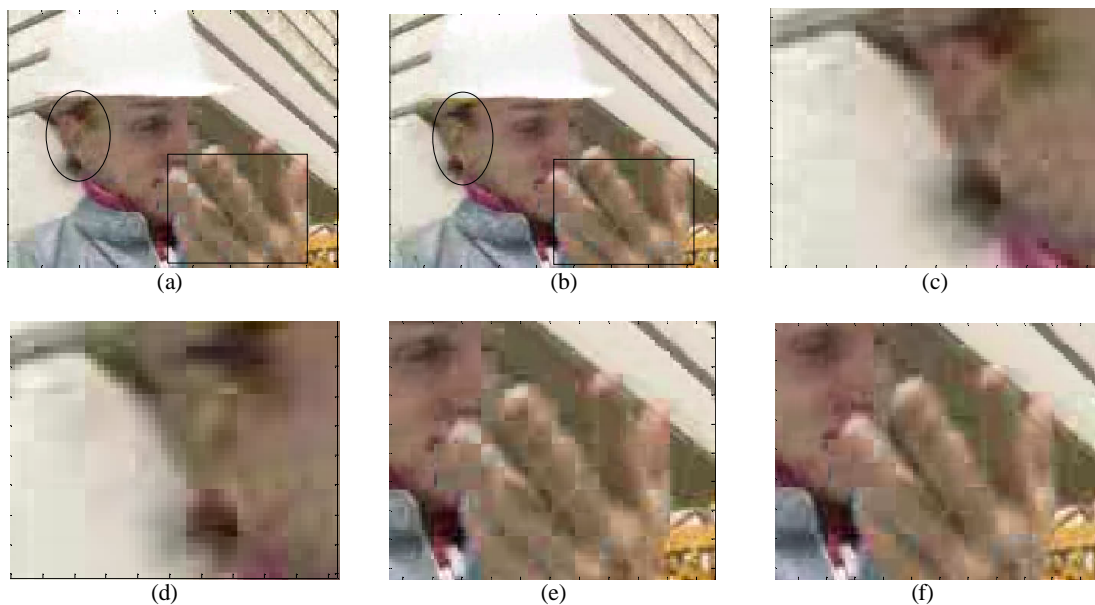


Fig.6. reconstructed frame 255, Left: Proposed; Right: Standard

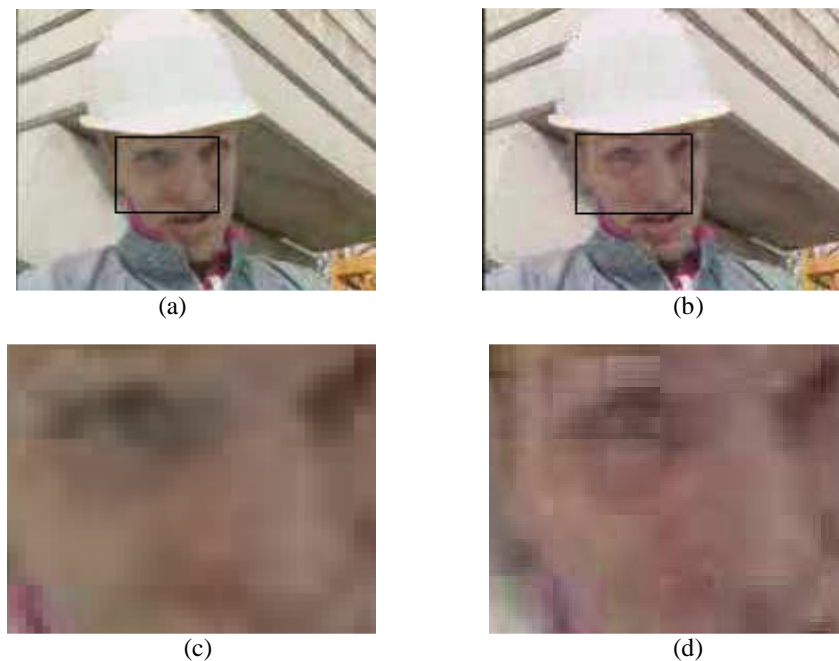


Fig.7. reconstructed frame 247, Left: Proposed; Right: H.263

## 4.2 Objective Evaluation

We also compared the objective video quality based on the same experimental conditions in section 4.1. The objective video quality is measured by Peak Signal to Noise Ratio (PSNR). According to the obtained results, the proposed model could achieve almost same objective quality as standard codec. Fig.8 gives the comparison between the proposed quantization method and the standard H.263 codec.

The experiments were carried out for different video sequences such as Coastguard, Suize and so on. The same evaluation results were reached. Furthermore, the utilization of the proposed method is also investigated in the experiments. Generally, the more motion activity the video sequence contains, the higher the utilization ratio of the proposed adaptive quantization approach. While for those sequences with no motion or only low local motion activity such as Trevor, the utilization of the proposed method is about 23%. Table II shows the utilization of the proposed scheme at the quantization parameter of 10. "Total" means the number of inter-coded DCT blocks in H.263 and "Proposed" is the number of blocks where HVS based quantization was applied.

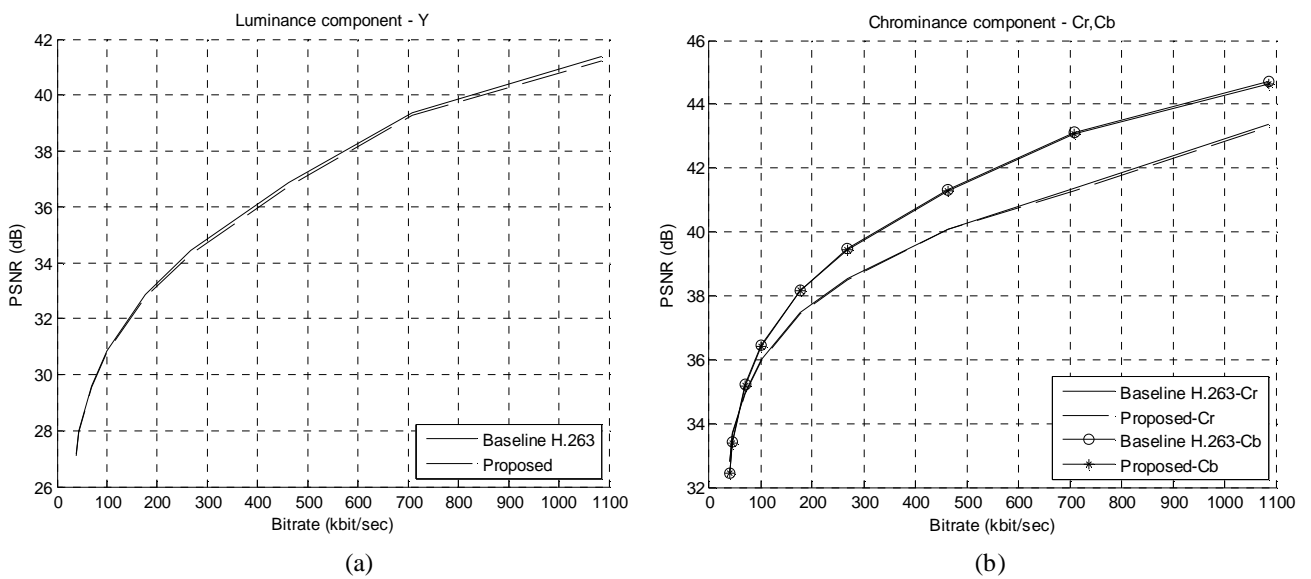


Fig. 8 Comparison of objective video quality between the proposed quantization algorithm and the baseline H.263 standard, where (a) shows PSNR vs. bitrate between the luminance components and (b) is the comparison of the chrominance components between the proposed method and the baseline H.263 codec.

TABLE II STATISTICAL RESULTS OF UTILIZATION

Sequence	Total	Proposed	Ratio
Foreman	395,364	202,966	51.31%
Coastguard	313,914	148,358	47.58%
Suzie	126,276	38,730	30.67%
Glasgow	733,530	381,342	51.99%
Trevor	132,078	30,228	22.89%
Carphone	358,872	129,372	36.05%

## 5. CONCLUSION

Block effect is an annoying artifact in digital video processing. Its influence on video objective quality is relatively slight. But it may degrade video subjective quality significantly. In this paper, we analyzed the HVS temporal characteristics in terms of spatial frequency and velocity. Furthermore, we proposed an adaptive inter quantization method by exploiting limitations of the human visual system. Experimental results show that, at the same average bitrate, the proposed model can improve the subjective video quality by decreasing visible block artifacts with nearly unchanged objective video quality. Best results can be expected for low-bitrate applications.

The proposed method need not change any bitstream grammar or syntax of the existing video coding standards. The modified encoder will still output standard-compliant bitstreams, which can be successfully decoded by any standard-compliant decoder. So it can be easily implemented on many quantization-based video codecs. The method is typically useful for digital applications at low bitrate communication such as video telephony and wireless communication.

## 6. ACKNOWLEDGEMENT

This work is supported partly by Chinese Science & Technology Ministry under Grant 2005DFA10300 and by Academy of Finland under Grant 117065.

## 7. REFERENCES

- [1] M. Ghanbari, "Video Coding: an introduction to standard codecs," The Institution of Electrical Engineers, London, United Kingdom, pp.98-99,1999.
- [2] W. Osberger, S. Hammond, and N. Bergmann, "An MPEG encoder incorporating perceptually based quantization," *IEEE, Speech and Imaging Technologies for Computing and Telecommunications*, pp. 731-733, 1997.
- [3] J. Malo, J. et. al., "Perceptual feedback in multigrid motion estimation using an improved DCT quantization", *IEEE Trans. on Image Processing*, 10(10) ,pp.1411-1427, 2001.
- [4] S. Zhou, J. Li, and Y. Zhang, "Efficient quantization step selection scheme for I-frame in rate constrained video coding," *IEEE ICIP*, pp.II-314-4, 2005.
- [5] H. Wang, M.Chan, S. Kwong, and chi-Wah Kok, "Novel quantized DCT for video encoder optimisation,"*IEEE Signal Processing Letters*, Vol.13, No.4, pp.205-208,2006.
- [6] B.Petljanski and O. Marques, "A novel approach for video quantisation using the spatiotemporal frequency characteristics of the human visual system," *British Machine Vision Conference*, 2005.
- [7] [online]www.fh-friedberg.de/fachbereiche/e2/telekomlabor/zinke/mk/mpeg2beg/whatisit.htm
- [8] D.H. Kelly, "Motion and vision ii. stabilized spatio-temporal threshold surface," *J.Opt. Soc. Am.*, pp.1340-1349, 1979.
- [9] D.W.Dong,"Spatiotemporal inseparability of natural images and visual sensitivities,"*Motion Vision-Computational, Neural, and Ecological Constraints*, New York, 2001
- [10] Recommendation ITU-R BT.500-11, "Methodology for the Subjective Assessment of the Quality of Television Pictures."