# Modeling of 3-D DCT Coefficients for Fast Video Encoding

Jin Li[1] Jarmo Takala[1]
Department of Information Technology
Tampere University of Technology
Tampere, Finland[1]
FirstName.LastName@tut.fi[1]

Moncef Gabbouj[1] Hexin Chen[2]
School of Communication Engineering
Jilin University
Changchun, China[2]
chx@email.jlu.edu.cn[2]

*Abstract* - **The discrete cosine transform (DCT) plays a key role in image/video coding due to its powerful energy concentration ability. On the other hand, since many DCT coefficients will be quantized to zeros, a huge number of redundant computations are introduced. This paper proposes an analytical model to predict the zero-quantized DCT (ZQDCT) coefficients for fast 3-D DCT based video encoding. The theoretical analysis is first performed to study the sufficient condition under which each DCT coefficient becomes zero. Then, the threshold for each ZQDCT coefficient is mathematically determined. Finally the transform and quantization operations of the predicted ZQDCT coefficients are omitted. We show, by the experimental results, that the proposed model can significantly reduce the redundant computations and achieve better real-time performance than the original encoder without any quality degradation. Computational reduction also implies longer battery lifetime and energy economy for digital applications.**

***Keywords- discrete cosine transform (DCT); quantization; video coding; computational complexity***

## I. INTRODUCTION

As an alternative approach to motion compensation in video coding standards (e.g., MPEG-1/2/4 and H.261/3), three-dimensional (3-D) DCT has been investigated by many researchers. The main advantage of 3-D DCT is the low computational complexity. This is very important for many real-time applications and digital devices as they are still suffering most from the restriction of battery lifetime. Although there is significant research in reducing the computations for H.264 and MPEG-4 [1]-[5], the complexity is still far from the requirements of most portable devices. Therefore, there is a great need now to develop sufficient 3-D DCT algorithms for such applications.

Previously, the efforts to advance the 3-D DCT algorithms are mainly focused on the improvement of coding efficiency. These proposed approaches can reduce the correlations more effectively by using the scene change detector [6], threshold for different local activity [7]-[8] and multiple description method [9]. The latest results [10] indicate that 3-D DCT coding outperforms MPEG-2 while is slightly worse than MPEG-4 at high bit rates.

Although 3-D DCT coding is much superior to H.264 and MPEG-4 in terms of computational complexity, it is still desired to further reduce the computations without visual quality degradation. As the most complex part, three-dimensional DCT transform consumes more than half of the computation in 3-D DCT encoding while a large number of transformed coefficients will be finally quantized to zeros. Thus, if we can reduce these redundant computations, the encoding process will be much faster. This also implies longer battery lifetime for portable digital applications.

In this paper, we propose an analytical model to skip redundant DCT and quantization without visual quality degradation through a theoretical analysis on the dynamic range of DCT coefficients. Although the proposed model is implemented based on the baseline 3-D DCT video coding, it can be widely used on other 3-D DCT schemes in [4]-[8]. As a result, high prediction efficiency and significant computational savings are achieved by the proposed model.

The rest of this paper is organized as follows. The threshold for each ZQDCT coefficient is mathematically analyzed and proposed in Section 2. The experimental results are presented in Section 3. Finally, Section 4 concludes this paper.

## II. PROPOSED ANALYTICAL MODEL

### A. Analysis of zero-quantized 3-D DCT coefficients

We firstly analyze the sufficient condition for quantized 3-D DCT coefficients to be zeros. In this paper, we mainly consider $8 \times 8 \times 8$ DCT which is widely used in many 3D DCT video coding. If we define $f(x, y, z)$ as the pixel value, $0 \leq x, y, z \leq 7$, the DCT coefficient $F(u, v, w)$, $0 \leq u, v, w \leq 7$, is computed by

$$F(u,v,w) = \frac{c(u)c(v)c(w)}{8} \sum_{x=0}^{7} \sum_{y=0}^{7} \sum_{z=0}^{7} f(x,y,z) \cos\frac{(2x+1)u\pi}{16} \times \cos\frac{(2y+1)v\pi}{16} \cos\frac{(2z+1)w\pi}{16} \quad (1)$$

TABLE I THRESHOLDS OF ZERO QUANTIZED 1-D DCT COEFFICIENTS ($0 \le y, z \le 7$)

| Threshold | DCT Coefficient $(u, y, z)$ |
|---|---|
| $T_1(y,z) = \frac{2\alpha}{\cos(\pi/16)}$ | $u = 1,3,5,7$ |
| $T_2(y,z) = \frac{2\alpha}{\cos(\pi/8)}$ | $u = 2,6$ |
| $T_3(y,z) = 2\sqrt{2}\alpha$ | $u = 4$ |

where $c(u)$, $c(v), c(w) = 1/\sqrt{2}$, for $u, v, w = 0$, and $c(u)$, $c(v), c(w) = 1$, otherwise. Then, the DCT coefficient will be quantized to zero if such condition holds true

$$F(u,v,w) < Q(u,v,w) \tag{2}$$

where $Q(u,v,w)$ is the quantization parameter for $F(u,v,w)$.

### B. Proposed model for zero-quantized prediction

Since the 3-D DCT can be calculated in a row-column-frame order, we first consider the row-wise 1-D DCT. If $\bar{f}(y,z)$ is the mean value of the eight pixels in each row and $f^r(x,y,z)$ is the residual pixel value, we define

$$\bar{f}(y,z) = \tfrac{1}{8}\textstyle\sum_{x=0}^{7} f(x,y,z), \quad f^r(x,y,z) = f(x,y,z) - \bar{f}(y,z) \tag{3}$$

$$\forall y,z \in \{0,1,\dots,7\}$$

Then, each DCT coefficient can be computed by the mean value $\bar{f}(y,z)$ and the eight residual pixel values $f^r(x,y,z)$ as

$$F(u,y,z) = \begin{cases} 2\sqrt{2}\bar{f}(y,z) & for \quad u = 0 \\ \frac{c(u)}{2}\sum_{x=0}^{7} f^r(x,y,z)\cos\frac{(2x+1)u\pi}{16} & otherwise \end{cases} \tag{4}$$

where $\forall y,z \in \{0,1,\dots,7\}$ and (5) gives the proof process.

For any quantization $Q(u,v,w)$, we can decompose it into the following format

$$Q(u,v,w) \approx \alpha \times \beta \times Q^f(u,v,w) \tag{6}$$

if we define

$$Q^f(u,v,w) = \left\lfloor \frac{Q(u,v,w)}{\alpha\beta} \right\rfloor \tag{7}$$

$$\forall u,v,w \in \{0,1,\dots,7\}$$

where $\alpha, \beta, Q^f(u,v,w)$ are the quantization parameters used to quantize the DCT coefficients after the transform in each stage in the row-column-frame order. And $\lfloor x \rfloor$ denotes the nearest integer less than or equal to $x$.

TABLE II THRESHOLDS OF ZERO-QUANTIZED 3-D DCT COEFFICIENTS ($0 \le u, v \le 7$)

| Threshold | DCT Coefficient $(u, v, w)$ |
|---|---|
| $T_1(u,v) = \frac{2Q^f(u,v,w)}{\cos(\pi/16)}$ | $w = 1,3,5,7$ |
| $T_2(u,v) = \frac{2Q^f(u,v,w)}{\cos(\pi/8)}$ | $w = 2,6$ |
| $T_3(u,v) = 2\sqrt{2}Q^f(u,v,w)$ | $w = 4$ |

Additionally, the sum of absolute difference $SAD(y,z)$ of the eight residual pixels in each row is given by

$$SAD(y,z) = \sum_{x=0}^{7} |f^r(x,y,z)| \tag{8}$$

$$\forall y,z \in \{0,1,\dots,7\}$$

From (4) and (8), the 1-D DCT coefficient $F(u,y,z)$ is bounded by

$$F(u,y,z) \le \frac{c(u)}{2} max\left\{\left|cos\frac{(2x+1)u\pi}{16}\right|\right\} \times SAD(y,z) \tag{9}$$

So, $F(u,y,z)$ can be predicted as zero if

$$SAD(y,z) \le \frac{2\alpha}{c(u) max\left\{\left|cos\frac{(2x+1)u\pi}{16}\right|\right\}} \tag{10}$$

$$for\ u \ne 0 \quad \forall y,z \in \{0,1,\dots,7\}$$

Therefore, we can predict $F(u,y,z)$ as zero by comparing $SAD(y,z)$ with the threshold in (9). Each DCT coefficient is bounded depending on the frequency position that affects the maximum value of the cosine function. As a result, the thresholds to determine zero quantized DCT coefficients are listed in Table I.

Similarly, we continue decompose $F(u,y,z)$ into a series of mean values $\bar{F}(u,z)$ and a residual cube $F^c(u,y,z)$ along column direction. The 2-D DCT coefficients $F(u,v,z)$ after the column transform will be predicted as zero if the following condition holds true

$$SAD(u,z) \le \frac{2\beta}{c(v) max\left\{\left|cos\frac{(2y+1)v\pi}{16}\right|\right\}} \tag{11}$$

$$for\ v \ne 0 \quad \forall u,z \in \{0,1,\dots,7\}$$

The thresholds for predicted zero-value coefficients are the same as in Table I just by replacing the quantization parameter $\alpha$ by $\beta$.

Finally, the 3-D DCT coefficient will be predicted as zero if

$$SAD(u,v) \le \frac{2Q^f(u,v,w)}{c(w)max\left\{\left|cos\frac{(2z+1)w\pi}{16}\right|\right\}} \quad (12)$$

$$for\ w \ne 0 \qquad \forall u,v \in \{0,1,\dots,7\}$$

and the thresholds for 3-D DCT coefficients to be quantized to zeros are listed in Table II.

Theoretically, the DCT coefficients can be most likely predicted as zeros in the following two situations: One, if all the eight values are very close to zeros (e.g. high frequency coefficients) and two, the variation among these values is small enough. Fig. 1 gives an example based on Akiyo sequence. Eight 2-D DCT coefficients after the row-column transform at $u,v = 0$ are shown in (a). Although these coeffcients are large, the residuals are small enough. (b) shows the eight high frequency coefficients at $u,v = 7$ and they are similar to the residual values in (a). Therefore, all the coefficients will be directly predicted as zeros without taking the temporal transform.

### C. Implementation of proposed analytical model

Based on the thresholds in Table I and II, we propose an algorithm to perform the 3-D DCT computations in the row-column-frame order. Table III shows the implementation of the row and column transform. Take the row-wise DCT for example, if $SAD(y,z) \le T_1(y,z)$, we only compute the first coefficient. If $(y,z) \le SAD(y,z) \le T_2(y,z)$ we only perform the 0, 1, 3, 5, 7 transform and quantization and skip the 2, 4, 6 coefficients. Otherwise if $SAD(y,z) \ge T_3(y,z)$, all transform and quantization are required to be calculated. For the frame transform, since the quantization $Q^f(u,v,w)$ is usually non-uniform, each coefficient has to be compared with its own threshold to decide to skip the transform or not.

As for the 3-D DCT implementation, we utilize the row-column-frame approach and butterfly-flow structure. Since we can predict some DCT coefficients as zeros in advance, the DCT computations of zero-valued coefficients can be skipped. For quantization, we just omit this step if the coefficients can be directly set to zeros.

## II. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed algorithm, a series of experiments were carried out with the baseline 3-D DCT codec. Four benchmark QCIF $(176 \times 144)$ sequences are tested. All the simulations are running on a PC with Intel Pentium 3.2G and 1.5Gbytes of RAM. The quantization strategy in accordance with [10] is defined as

$$q\overrightarrow{[k]} = \lfloor Q_p(1 + k_1^p + k_2^p + k_3^p) \rfloor \quad (13)$$

where $q\overrightarrow{[k]}$ is the quantization step for 3-D DCT coefficient at position $\vec{k}[k_1,k_2,k_3]$, $Q_p$is the quantization parameter, and $p$ is the parameter to control the rate of delay in the quantization

| | 239 | 238 | 239 | 239 | 239 | 239 | 240 | 238 |
|---|---|---|---|---|---|---|---|---|
| **239** | **0** | **-1** | 0 | 0 | 0 | 0 | 1 | -1 |

(a)

| | 0 | 0 | 0 | -1 | 1 | 0 | 0 | 1 |
|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 0 | 0 | -1 | 1 | 0 | 0 | 1 |

(b)

Figure. 1 Example of DCT coefficients to be predicted as zeros

TABLE III IMPLEMENTATION OF PROPOSED ANALYTICAL MODEL ON EACH STAGE

| Type | Condition | Strategy to implementation |
|---|---|---|
| 1 | $SAD \le T_1$ | only the first coefficient |
| 2 | $T_1 \le SAD \le T_2$ | only coefficients 0,1,3,5,7 |
| 3 | $T_2 \le SAD \le T_3$ | only coefficients 0,1,2,3,5,6,7 |
| 4 | $T_3 \le SAD$ | all coefficients |

volume. Here, it is fixed to 0.3. In addition, the parameters $\alpha,\beta$ in (6) are 2 throughout the whole experiments. As $\alpha,\beta$ are small enough, it doesn't cause any information loss compared to the standard DCT and quantization approach in the baseline 3-D DCT codec.

### A. Computational reduction of DCT and quantization

Firstly, we will study the computational complexity of the proposed analytical model. The comparisons of the computational complexity about 3-D DCT and quantization between the proposed model and the original encoder are illustrated in Fig.2. In this figure, the required computational complexity for the proposed model is defined as

$$C = \frac{T_d}{T_d^o} \times 100\% \quad (14)$$

where $T_d$ and $T_d^o$are the encoding time of DCT and quantization for the proposed model and the baseline codec. In the figure, the computational reduction in each stage (i.g., row, column and frame) is explicitly shown in (a)-(c). In addition, Fig.2 (d) gives the overall computational savings in 3-D DCT and quantization. It is obvious that the proposed model can effectively reduce redundant computations and achieve better performance than the original codec in terms of computational cost. In general, the average computations of 3-D DCT have been decreased by 50% compared to the original encoder, although the extent is slightly different for different video sequences and quantization parameters.

### B. False acceptance rate and false rejection rate

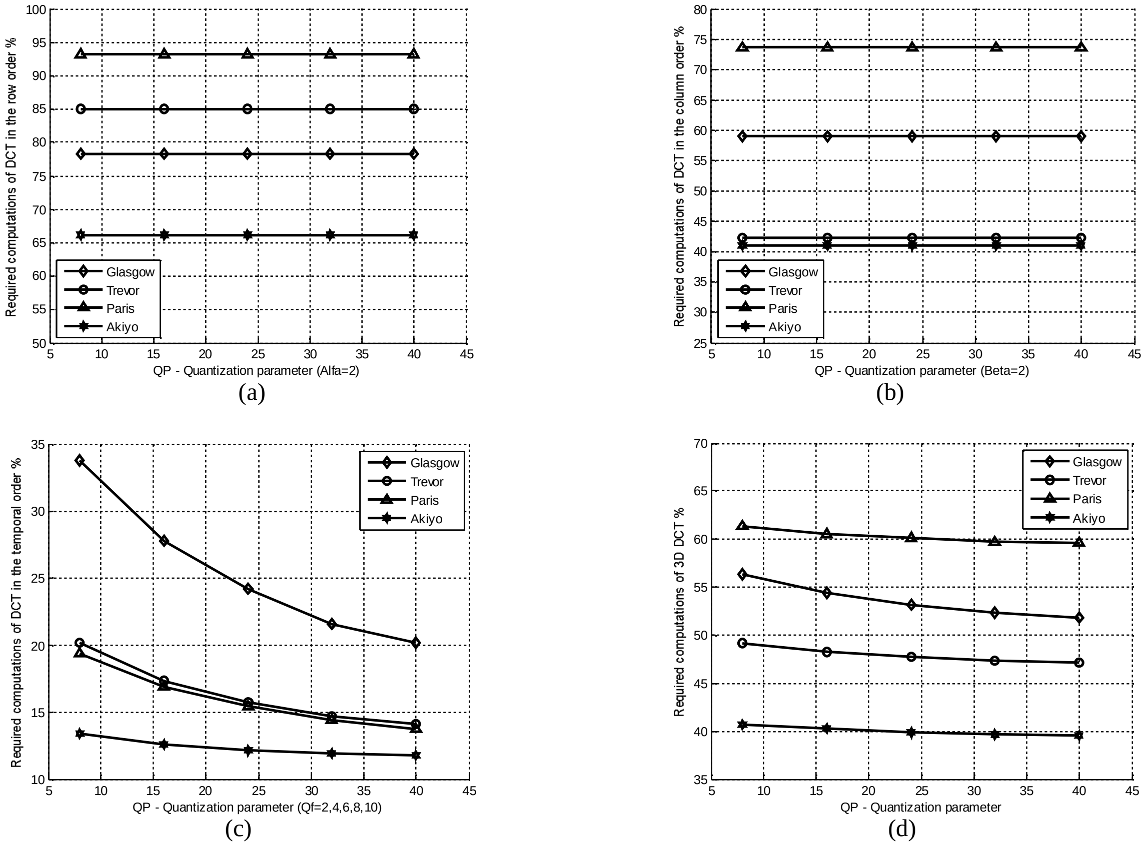


Figure.2 Required computations of 3-D DCT where (a), (b), (c) are the required computations in the row, column, frame-wise transform and (d) is the total required computations compared to the baseline 3-D DCT approach.

As two important evaluation parameters, the false acceptance rate (FAR) and the false rejection rate (FRR) are provided to evaluate the proposed analytical model. The FAR and FRR are defined as

$$FAR = \frac{N_{mn}}{N_n} \times 100\%, \quad FRR = \frac{N_{mz}}{N_z} \times 100\%, \qquad (15)$$

$N_{mn}$ is the number of non-zero quantized coefficients being miss classified as zero quantized DCT coefficients, $N_n$ is the total number of non-zero quantized coefficients. While $N_{mz}$ is the number of zero quantized coefficients being miss classified and $N_z$ is the total number of zero quantized coefficients. The smaller the FAR, the less the video quality degrades and the smaller the FRR, the more efficient the predictive model. Therefore, it is desirable to have both small FAR and FRR for an efficient predictive model.

From the experimental results, the proposed model has a zero FAR, which means that there is no false acceptance of zero quantized coefficients as we expected since the proposed model is derived based on mathematically verified analysis. So we only list the FRR results in Table IV for different video sequences. Averagely, the proposed model can predict about 70% of zero quantized coefficients.

### *C. Video quality and encoding time comparison*

Finally, we will study the video quality and the encoding time of the proposed model compare to the original codec. The objective video quality is measured by the Peak Signal to Noise Ratio (PSNR), no visual degradation is observed for the proposed model. This is exactly in accordance with FAR. Table V shows the entire encoding time of the proposed model, where the encoding time reduction $\nabla T$ is presented as

$$\nabla T = \frac{T}{T_{org}} \times 100\% \qquad (16)$$

where $T_{org}$ and $T$ are the entire encoding time of the baseline 3-D DCT encoder and the proposed model.

From Table V, it is obvious that the real time performance based on our analytical model is much better than the original codec. This validates that the proposed model can reduce the computational complexity of the video encoder and is more practical for real-time applications and portable digital devices.

## III. CONCLUSION

This paper proposes an analytical model to predict zero-quantized 3-D DCT coefficients to avoid redundant DCT and quantization computations. We analyzed zero-quantized DCT coefficients and derived a sufficient condition under which

TABLE IV STATISTICAL RESULTS OF FRR (%)

| Stage | $Q_p$ | Akiyo | Trevor | Paris | Glasgow |
|---|---|---|---|---|---|
| Row | $\alpha = 2$ | 39.20 | 65.23 | 70.40 | 54.73 |
| Column | $\beta = 2$ | 27.27 | 25.68 | 47.27 | 33.64 |
| Frame | 2 | 9.61 | 16.81 | 8.05 | 27.89 |
| | 4 | 10.80 | 15.82 | 11.78 | 24.84 |
| | 6 | 11.10 | 14.85 | 12.46 | 22.17 |
| | 8 | 12.20 | 14.07 | 12.23 | 19.98 |
| | 10 | 10.96 | 13.15 | 11.91 | 18.32 |
| Total | 8 | 22.85 | 29.65 | 31.29 | 36.00 |
| | 16 | 23.22 | 29.07 | 32.42 | 34.43 |
| | 24 | 23.31 | 28.59 | 32.51 | 33.14 |
| | 32 | 23.33 | 28.22 | 32.38 | 32.08 |
| | 40 | 23.29 | 27.84 | 32.21 | 31.16 |

each DCT coefficient becomes zero. Based on the analysis, an efficient detection algorithm is proposed. The experimental results show that the proposed model can significantly improve the encoding efficiency without visual quality degradation. Moreover, it can be directly applied to any other existing 3-D DCT schemes. Computational reduction also implies longer battery lifetime and energy economy for digital applications. Potential applications could be for portable digital devices with restrict battery lifetime and other areas with real-time requirement.

## IV. ACKNOWLEDGMENT

This work is supported partly by Chinese Science & Technology Ministry under Grant 2005DFA10300 and by Academy of Finland under Grant 117065. The authors would like to thank Nokia Foundation for the support of research work.

## REFERENCES

[1] I.M. Pao and M.T. Sun, "Modeling DCT Coefficients for Fast Video Encoding," IEEE *Transactions on Circuit and Systems for Video Technology*, VOL 9, No. 4, pp. 608-616, 1999.


TABLE V COMPARISON OF ENCODING TIME (%)

| $Q_p$ | Akyio | Trevor | Glasgow | Paris |
|---|---|---|---|---|
| 8 | 50.40 | 62.37 | 70.85 | 77.74 |
| 16 | 49.29 | 60.23 | 67.87 | 75.50 |
| 24 | 48.98 | 59.31 | 66.21 | 74.89 |
| 32 | 48.77 | 58.86 | 65.09 | 74.37 |
| 40 | 48.62 | 58.57 | 64.26 | 73.83 |

[2] Y.H. Moon, G.Y. Kin and C.W. Kok, "An Improved Early Detection Algorithm for All-Zero Blocks in H.264 Video Encoding," IEEE *Transactions on Circuit and Systems for Video Technology*, VOL 15, No. 8, pp. 1053-1057, 2005.

[3] H.L. Wang, S. Kwong and C.W. Kok, "Efficient Prediction Algorithm of Integer DCT Coefficients for H.264/AVC Optimization," IEEE *Transactions on Circuit and Systems for Video Technology*, VOL 16, No. 4, pp. 547-552, 2006.

[4] A. Docef, F. Kossentini, K. Nguuyen-Phi and I. R. Ismaeil, "The Quantized DCT and Its Application to DCT-Based Video Coding," IEEE *Transactions on Image Processing*, VOL 11, No. 3, pp. 177-187, 2002.

[5] H.L. Wang, S. Kwong and W.C. Siu, "Analytical Model of Zero Quantized DCT Coefficients for Video Encoder Optimization," IEEE *Proceedings on ICME*, pp. 801-804, 2006

[6] Y.L. Chan and W.C. Siu, "Variable Temporal-length 3-D Discrete Cosine Transform Coding," IEEE *Transactions on Image Processing*, VOL 6, No. 5, pp. 758-763, 1997.

[7] N.P. Sgouros, S.S. Athineos, P.E. Mardaki and *etc.*, "Use of an Adaptive 3D-DCT Scheme for Coding Multiview Stereo Images," IEEE *Proceedings of ISSPIT*, pp.180-185, 2005.

[8] B. Furht, K. Gustafson, H. Huang, and O. Marques, "An Adaptive Three-Dimensional DCT Compression Based on Motion Analysis," *Proceedings of ACM*, pp. 765-768, 2003.

[9] A. Nokin, A. Gotchev, K. Egiazarian and J. Astola, "A Low-Complexity Multiple Description Video Coder Based on 3D-Transforms," EURASIP *Journal on Embedded Systems*, 2007.

[10] N. Bozinovic, and J. Konrad, "Motion Analysis in 3D DCT Domain and Its Application to Video Coding," EURASIP *Signal Process., Image Commun*., June,2005.

$$F(u,y,z) = \frac{c(u)}{2}\sum_{x=0}^{7} f(x,y,z)cos\frac{(2x+1)u\pi}{16} = \frac{c(u)}{2}\sum_{x=0}^{7}\left[f^{r}(x,y,z)+\bar{f}(y,z)\right]cos\frac{(2x+1)u\pi}{16}$$

$$= \frac{c(u)}{2}\sum_{x=0}^{7} f^{r}(x,y,z)cos\frac{(2x+1)u\pi}{16} + \frac{c(u)}{2}\bar{f}(y,z)\sum_{x=0}^{7} cos\frac{(2x+1)u\pi}{16} \quad (5)$$

Since

$$\sum_{x=0}^{7} f^{r}(x,y,z) = 0 \quad for\ u = 0 \qquad and \qquad \sum_{x=0}^{7} cos\frac{(2x+1)u\pi}{16} = 0 \quad for\ u \neq 0$$

Thus, (4) is verified.