Predictive Orthogonal Projection for Low-Complexity Multi-view Video Coding

Zhenglin Wang
School of Computer and Information Science
The University of South Australia
Adelaide, Australia
wanzy047@mymail.unisa.edu.au

Ivan Lee
School of Computer and Information Science
The University of South Australia
Adelaide, Australia
Ivan.Lee@unisa.edu.au

Abstract—**Two-Dimensional Discrete Cosine Transform (2-D DCT) has been playing an important role in modern image and video codecs. However, the computational complexity of 2-D DCT might be still a limitation for many emerging applications such as Multi-view Video Coding (MVC). In this paper, we propose a novel transform of predictive orthogonal projection (POP) for low-complexity MVC. POP firstly trains a subset of video blocks of a multi-view video sequence to generate adaptive transform matrices. Then, these adaptive transform matrices are used to transform all video blocks in the video sequence. Compared to 2-D DCT, POP is 1-dimensional transform. If the length of the original signal is N and M (M<N) transform coefficients are preserved for reconstruction, the complexity of 2-D DCT is usually O(N2) while that of POP is O(MN). So, when high compression ratio is desirable, POP will outperform 2-D DCT in computational cost. Experimental results also demonstrate that POP exhibits better computational performance than 2-D DCT without degrading the rate-distortion performance when raw video frames are divided into medium sized blocks.**

Keywords-multi-view video coding (MVC); random projection (RP); 2-D DCT; clustering; principal component analysis (PCA)

I. INTRODUCTION

Random projection (RP) has attracted great interest in dimension reduction on image data because of its low computational complexity [1]. However, RP underperforms 2-D DCT in terms of image compression [1]. Transform-based coding such as 2-D DCT is widely adopted in popular image and video compression standards, such as JPEG and H.264. The essence of 2-D DCT is that it can concentrate most of natural image signals information in a few low-frequency DCT coefficients by using an invariable and predefined transform matrix [2]. Therefore, 2-D DCT based compression algorithms can achieve high compression ratio (1-D DCT is weak in image and video compression compared with 2-D DCT). Since 2-D DCT is 2-dimensional transform while RP is 1-dimensional operation, RP usually outperforms 2-D DCT if only regarding computational complexity [1].

Recently, Multi-view video coding (MVC) is recognized to be one of the key technologies for a wide variety of applications such as free viewpoint video (FVV), 3DTV, distributed video sensor networks [8] and multi-camera video surveillance. In those applications, the multi-view video signals usually comprise huge amount of image data so both computational complexity and compression ratio are crucial for MVC. This paper proposes a novel transform of predictive orthogonal projection (POP) to reduce the complexity of MVC. Similar to RP, POP is a 1-dimensional transform. But, RP projects a high-dimensional image signal onto a random low-dimensional subspace while POP projects it onto a trained low-dimensional subspace. The trained subspace for POP is abstracted from partial data of the target signal; therefore, the target signal can probably obtain an orthogonal projection which highly approximates itself. In MVC, since the similarities exist in temporally and spatially adjacent frames, it is worthwhile to train part of them to obtain an optimized subspace, which is also the row space of the transform matrix. When the dimension of the trained subspace is smaller than that of the original signal, compression will be built in the POP transform. The novel feature of POP transform over 2-D DCT is helpful to reduce the encoding complexity. The recovery process of POP is to retrieve the related orthogonal projection.

The underlying concept of POP is to explore adaptive transform matrices for different video signals. On the contrary, the popular 2-D DCT maintains an invariable transform matrix for arbitrary visual signals, and 2-D DCT is used as a benchmark in this paper. Since this paper focuses on low-complexity encoding and decoding, complex compression algorithms such as motion estimation are not considered. Clustering and principal component analysis (PCA) are common data mining techniques [5][6]. Clustering is computationally efficient while principal component analysis (PCA) yields the optimal orthogonal transform matrix [6], so both techniques are used for POP to train the adaptive transform matrices from the training data.

II. PREDICTIVE ORTHOGONAL PROJECTION

RP has been found to be a computationally efficient method for dimension reduction of high-dimensional data sets. In RP, the original set of K n-dimensional observations denoted by $\mathrm{X}_{\mathrm{n}\times\mathrm{K}}$, is projected to an m-dimensional (m ≪ n) subspace with using a random matrix $\mathrm{A}_{\mathrm{m}\times\mathrm{n}}$, and formed into a set of K m-dimensional projections $\mathrm{Y}_{\mathrm{m}\times\mathrm{K}}$. RP is usually presented as:

$$\mathrm{Y}_{\mathrm{m}\times\mathrm{K}} = \mathrm{A}_{\mathrm{m}\times\mathrm{n}}\mathrm{X}_{\mathrm{n}\times\mathrm{K}} \tag{1}$$

The key idea of RP is that if points in a vector space are projected onto a randomly selected subspace of an adequate dimension, the distances between these points will be approximately preserved [1].

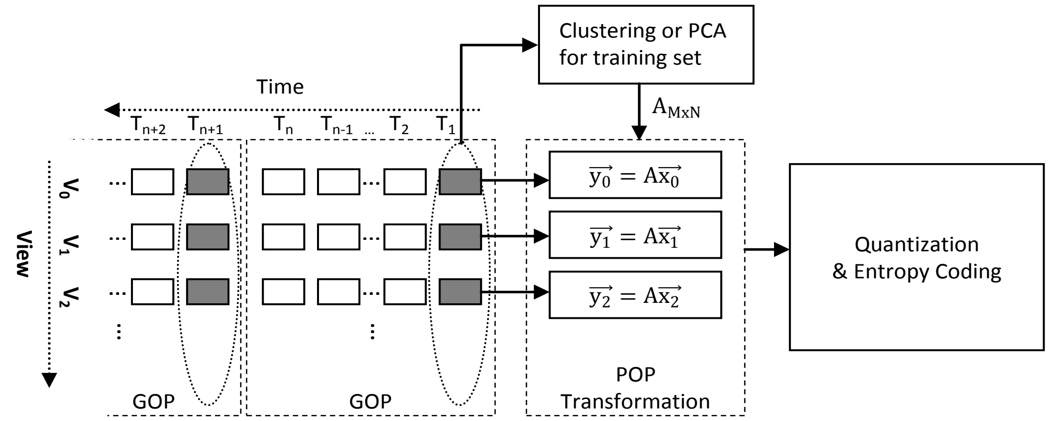

Clustering or PCA for training set
Time
A_{MxN}
T_{n+2}
T_{n+1}
T_n
T_{n-1}
...
T_2
T_1
View
V_0
V_1
V_2
$\overrightarrow{y_0} = A\overrightarrow{x_0}$
$\overrightarrow{y_1} = A\overrightarrow{x_1}$
$\overrightarrow{y_2} = A\overrightarrow{x_2}$
Quantization & Entropy Coding
GOP
GOP
POP Transformation

Figure 1. Low-complexity MVC based on POP transform

The nature of RP suffers from the following two issues: 1) the projection points approximately preserve the relative distances between the original points, but they are probably quite different from the original points; 2) the column vectors of $Y_{m\times K}$ are not real projections of $X_{n\times K}$ [1]. The proposed POP technique can overcome these issues.

Let $\vec{x} = [x_1, x_2, \cdots, x_n]^T$ denote a vectorized n-pixel image signal, $A_{mxn} = (\vec{a}_1|\vec{a}_2| \dots |\vec{a}_m)^T$ denote an m-by-n transform matrix and vectors $(\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m)$ are linearly independent, then the orthogonal projection $\vec{p}$ [3] of $\vec{x}$ onto $H = \mathrm{range}(A^T)$ is calculated as (the subscripts for dimensions are omitted):

$$\vec{p} = A^T(AA^T)^{-1}A\vec{x} \tag{2}$$

and equivalently,

$$A\vec{p} = A(A^T(AA^T)^{-1}A\vec{x}) = A\vec{x} \tag{3}$$

If $A\vec{p}$ is notated by $\vec{y}$, (3) can be rewritten as:

$$\vec{y} = A\vec{x} \tag{4}$$

Equation (4) is the proposed POP transform. When $K = 1$, equation (1) is identical to equation (4); thus POP inherits the advantage of high computational efficiency.

In this paper, POP is not designed for dimension reduction; rather, it is proposed for image and video compression. It is known that the best approximation of a vector in a space is its orthogonal projection onto this space [3]. Naturally, the reconstruction of POP is to retrieve the orthogonal projection of $\vec{x}$ in the space H.

Since $\vec{p}$ is the orthogonal projection of $\vec{x}$ onto H, $\vec{p} \in H$, there must be a vector $\vec{w}_1 \in \mathbb{R}^m$ which satisfies $\vec{p} = A^T\vec{w}_1$. Then,

$$\vec{y} = A\vec{p} = A(A^T\vec{w}_1) \Rightarrow \vec{w}_1 = (AA^T)^{-1}\vec{y} \tag{5}$$

Subsequently, $\vec{p}$ can be solved as

$$\vec{p} = A^T\vec{w}_1 = A^T(AA^T)^{-1}\vec{y} \tag{6}$$

Equation (6) represents the POP reconstruction. Note that $\vec{p} \neq \vec{x}$ except $\vec{x} \in H$. Therefore, probably POP embodies a lossy compression when $m < n$.

If a random matrix is selected as the POP transform matrix, in other words, the original image signal is projected onto a random subspace, POP is identical to RP. Even if the orthogonal projection is solved by reconstruction, the reconstructed image might be significantly different to the original image and the loss of fidelity is not desirable. So, the selection of transform matrix is crucial to POP. The selected POP transform matrix is expected to be predictive and adaptive to different target signals so that the reconstruction process can retrieve a highly approximate orthogonal projection. In MVC, since similarities exist among neighboring frames both in temporal domain and spatial domain, it is possible to obtain such a predictive and adaptive transform matrix by training a subset of these frames.

III. MVC Based on POP

Low encoding complexity is significant for many emerging MVC applications such as distributed video sensor networks [8] and wireless multimedia sensor networks [10]. However, as multiple cameras are used to acquire the video signals at different view-angles, the amount of data to be processed increases tremendously [4]. Subsequent issues including compression ratio and encoding complexity are arising as well. Currently, most existing MVC techniques are based on variations of 2-D DCT, which suffer from the following two disadvantages: 1) 2-D DCT is more complex than the 1-dimensional transform such as RP. Furthermore, a zigzag scan is necessary to concentrate the important DCT coefficients into the lower order space before further compression, which also increases the computational cost; 2) The DCT matrix is invariable, so it is hardly optimal for all multi-view video signals.

In this paper, POP is proposed to be used for MVC to decrease the computational complexity. For multi-view video signals, high similarities exist among not only temporally consecutive frames but also spatially adjacent frames [4], so it is worthwhile to train a subset of these frames to obtain a preferable transform matrix for a cluster of similar frames. Clustering and PCA are common data analysis techniques to extract a feature subspace from high dimensional data sets. PCA is an effective approach in dimension reduction, but has high computational requirements compared with clustering. So, both these two solutions are investigated in this paper for

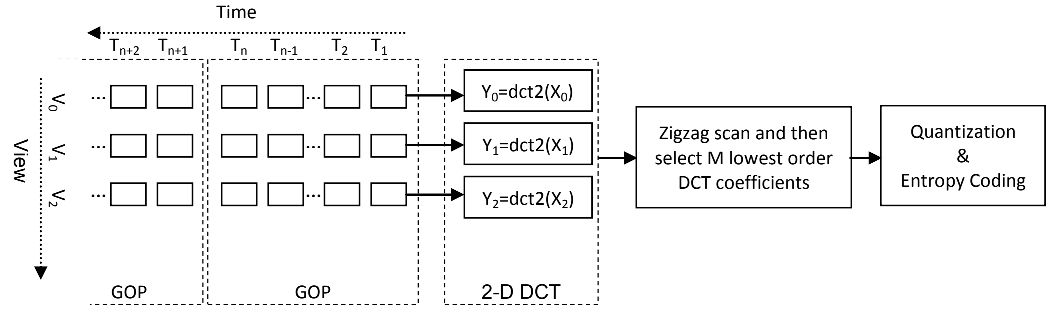

Time
T_{n+2} T_{n+1} T_n T_{n-1} T_2 T_1
View
V_0
V_1
V_2
Y_0=dct2(X_0)
Y_1=dct2(X_1)
Y_2=dct2(X_2)
Zigzag scan and then select M lowest order DCT coefficients
Quantization & Entropy Coding
GOP
GOP
2-D DCT

Figure 2. A comparison framework of MVC based on 2-D DCT

satisfying different applications: fast encoding or good reconstruction quality. Similar to 2-D DCT, block-based strategy is adopted in the proposed POP scheme due to the following two reasons: 1) if the dimension is excessively large, the high demand of memory and computational power is not desirable for low-complexity MVC applications; 2) because the transform matrices are variable, they are necessarily transmitted to the decoder along with the encoded transform coefficients. If the transform matrices are too large, they might produce a negative impact on compression ratio. We will present the effect of different block sizes in the following section.

The framework of POP-based MVC is illustrated in Figure 1. As the block-based strategy is employed, all frames are divided into equal-size, non-overlapping blocks. Then, all these blocks are vectorized according to equation (7) where X denotes a 2-dimensional n-by-n image block and $\vec{x}$ denotes the corresponding vector.

$$X = \begin{bmatrix} x_{11} & x_{12} & \vdots & x_{1n} \\ x_{21} & x_{22} & \vdots & x_{2n} \\ \cdots & \cdots & \vdots & \cdots \\ x_{n1} & x_{n2} & \vdots & x_{nn} \end{bmatrix},$$

$$\vec{x} = [x_{11}, \cdots, x_{n1}, x_{12}, \cdots x_{n2}, \cdots, x_{1n}, \cdots, x_{nn}]^T \quad (7)$$

Similar to MPEG video codec standards, Group of pictures (GOP) is used to express a group of successive temporal frames in a view in this paper. All views have the same GOP length. The GOPs of all views at the same time point are jointly compressed with using the POP scheme. The blocks of the first frames in a group of GOPs are used for training data besides being encoded. Each vectorized image block in the training set is an observation. Clustering or PCA is performed on the training set to learn an adaptive transform matrix $A_{M\times N}$ where $N = n^2$ and M is determined by the compression ratio $\frac{N}{M}$. For clustering, the training set is grouped into M clusters and the centroids of all clusters are organized into $A_{M\times N}$; K-means++ [7] is employed because it can improve both the speed and the accuracy in comparison with the conventional k-means method. For PCA, the M most significant principal components are organized into $A_{M\times N}$. Then, $A_{M\times N}$ is used to transform each image block in the group of GOPs to obtain M POP transform coefficients. Next, the POP transform coefficients can adopt conventional quantization and entropy coding schemes to further improve the compression ratio before they are transmitted to the decoder. The decompression algorithm of POP for the decoder is specified in equation (6). Finally, all reconstructed blocks are reassembled into reconstructed frames. When PCA is used for the training set, equation (6) can be simplified as equation (8). Because the principal components are orthonormalized, $AA^T = I$ where I is an M-by-M identity matrix.

$$\vec{p} = A^T\vec{y} \quad (8)$$

2-D DCT only transmits the important DCT coefficients to the decoder because of its universal and predefined transform matrix while POP has to transmit not only the transform coefficients but also the adaptive transform matrices. Therefore, for POP, the adaptive transform matrices are necessarily counted in the overall compression ratio. Note that all image blocks in the same time-point GOPs of all views use the same POP transform matrix learned by clustering or PCA. Equation (9) is used to measure the ratio between the size of the POP transform matrix and the size of all frames in a joint group of GOPs, denoted by "TMRatio" (only the luminance component is considered here). Respectively, W, H, nViews and nGOP denote the video frame width and height, the number of views and the GOP length.

$$\text{MRatio} = \frac{W \times H \times \text{nViews} \times \text{nGOP}}{M \times N} \quad (9)$$

Then, the compression ratio for 2-D DCT is $\frac{N}{M}$ while that for POP is $(\frac{N}{M} + \text{TMRatio})$. In this paper, the GOP length is referring to the MPEG standards and chosen as 15. If the GOP length is too big, the frames in the joint group of GOPs might be short of sufficient similarities; likewise, if the GOP length is too small, the compression ratio for POP might be unacceptable.

Finally, considering the computational complexity: the computational complexity of standard 2-D DCT without optimization is of the order $O(N^2)$; POP is comparable to RP with the order $O(MN)$ [1]. In most MVC applications, M is much less than N for the sake of high compression ratio, so POP can significantly reduce the computational cost compared with 2-D DCT. With respect to the training algorithms, K-means++ has the complexity of $O(MKN)$ [7], and PCA has the complexity of $O(KN^2) + O(N^3)$ [9] (K is the number of observations and generally $K > N$). K-means++ has much lower complexity than PCA when M≪N. On the other hand, although both clustering and PCA are more time-consuming than 2-D DCT, they only execute once for each joint group of GOPs. Therefore, POP is potential to facilitate the complexity of the MVC encoder. And the experiment in the next section will demonstrate its feasibility.

IV. Experimental Results

In this section, we illustrate the experimental results based on four standard multi-view video sequences: Ballroom, Vassar, Exit (all of them are 640x480, 8-view clips) and BMX (1920x1080, high-definition stereoscopic video). The experiment is implemented with Matlab and the simulation is run on ThinkPad x200 (Intel Core2 Duo CPU P8600 @ 2.40GHz 2.39Ghz, 1.94GB of RAM). Similar to most studies on video compression, only the luminance components are reported.

The proposed technique is benchmarked against 2-D DCT, and the 2-D DCT based encoding process is illustrated as Figure 2. 2-D DCT is directly applied to the image blocks (we directly call the function *dct2* provided by Matlab Image Processing Toolbox). Then, the DCT coefficients are scanned using a zigzag scan order. Finally, the M lowest order coefficients are selected and transmitted to the decoder (we do not select the M largest DCT coefficients to avoid recording the positions). Since we concentrate on comparing the performance between POP and 2-D DCT, conventional quantization and entropy coding are unemployed in our experiments. The compression ratio is the same for all compression schemes. In some test cases, M for 2-D DCT is bigger than M for POP because the POP transform matrix consumes a considerable part of compression ratio. To adapt 2-D DCT, all frames are divided into equal-size, non-overlapping and square blocks. Different block sizes are inspected in the experiment. The GOP length is 15 for each view. We only encode and decode the first GOPs of all views for each video sequence. To balance video quality and compression ratio, compression ratio is selected as 10:1.

Table-I, Table-II, Figure 3 and Figure 4 present the experimental results between POP with Clustering, POP with PCA and 2-D DCT. M is the number of selected encoded coefficients for each image block. Training time is the computational cost of learning the POP transform matrix, so the encoding time for the POP-based schemes includes the POP transform time and the training time. For the 2-D DCT, the encoding time only includes the 2-D DCT time. APSNR is the average PSNR for all the frames in a joint group of GOPs.

As described in Section III, clustering runs faster than PCA, and POP outperforms 2-D DCT in the time complexity. Therefore, POP with clustering should exhibit the fastest encoding speed. Practically, the experimental results in Table-I also demonstrate that POP with clustering has the best performance in terms of the encoding time. However, the reconstruction algorithm for POP with clustering (Equation (6)) is so complex that its decoding time is undesirable. On the other hand, the reconstruction algorithm for POP with PCA (Equation (8)) is simplified so that its decoding speed is much faster. The experimental results reveal that, when the block size is moderate such as 16-by-16 and 20-by-20, POP with PCA can achieve good performance in both the encoding and decoding time, and at least 60% improvement in comparison with 2-D DCT/IDCT. Choosing a moderate block size is practical. Small block size is disadvantageous to batch process so that the encoding and decoding time is big for both POP transform and 2-D DCT. Large block size potentially leads to high demand on memory and computational ability. In conclusion, POP with clustering is useful for those applications which require fast encoding; POP with PCA is promising to facilitate low-complexity MVC by considering both the encoding and decoding aspects.

Then, the reconstruction quality between the above three schemes is observed. Table-II records the PSNR performance of the POP-based schemes compared with the 2-D DCT at different block size. Figure 3 shows the frame-by-frame PSNR comparison and Figure 4 illustrates the comparison of visual qualities (Only some test results for Ballroom and BMX at medium block size are demonstrated due to the limitation of space; Ballroom presents the worst case and BMX presents the best case). The experimental results show that POP with

TABLE I. COMPUTATIONAL COST COMPARISON (COMPRESSION RATIO IS 10:1, ENCODING 15 FRAMES PER VIEW)

Video Name	Block size	POP					2-D DCT		
		M	Clustering		PCA		M	Enc. Time (s)	Dec. Time (s)
			Enc. Time (s) (inc. Training Time)	Dec. Time (s)	Enc. Time (s) (inc. Training Time)	Dec. Time (s)			
Ballroom (640x480, 8 views)	8x8	6	5.13	32.65	5.51	14.68	6	66.65	85.18
	16x16	26	2.42	37.05	**3.46**	**6.38**	26	20.41	26.42
	32x32	102	3.13	222.34	54.36	6.22	105	8.15	11.52
Vassar (640x480, 8 views)	8x8	6	5.02	32.85	5.30	14.76	6	65.49	86.34
	16x16	26	2.30	44.61	**3.38**	**6.41**	26	19.56	26.38
	32x32	102	2.80	207.86	59.68	6.08	105	9.22	11.55
Exit (640x480, 8 views)	8x8	6	4.72	32.40	5.30	15.38	6	66.46	87.10
	16x16	26	2.20	37.91	**3.36**	**6.37**	26	19.55	26.42
	32x32	102	2.99	211.55	59.22	6.12	105	8.03	11.58
BMX (1920x1080, 2 views)	8x8	6	8.04	53.86	8.73	24.09	6	113.77	141.20
	20x20	40	3.63	81.27	**6.03**	**9.54**	40	24.32	32.22
	40x40	160	6.41	670.02	63.81	12.54	167	10.75	17.48

clustering performs worse in reconstruction quality than the other two schemes though it has the advantage of fast encoding speed. POP with PCA achieves the highest APSNR in all test cases. It should be pointed out that some individual frame PSNRs for 2-D DCT are higher than those for POP with PCA when the block size is 8-by-8, 32-by-32 or 40-by-40. The potential reasons are as below: 1) Similar to RP for dimension reduction, POP might be unsuitable for low-dimensional signals (small block size); 2) When block size is big, the predefined DCT transform matrix is advantageous to improve compression ratio while the large adaptive POP transform matrices may produce a negative impact on compression ratio. However, when the block size is 16-by-16 or 20-by-20, POP with PCA outperforms 2-D DCT in both average PSNR (Table-II) and per frame PSRN (e.g. Figure 3) on all test video sequences. Although minor PSNR improvements are achieved in Ballroom and Vassar, Exit and BMX videos show noticeable improvements with PSNR gains up to 0.89 dB. In addition, a perceptible visual quality improvement can be observed comparing POP with PCA (Figure 4-f) with 2-D DCT (Figure 4-g). The stripes on the rider and bicycle in Figure 4-f are clear and obvious, but they are blurring in Figure 4-g.

TABLE II. VIDEO QUALITY COMPARISON IN PSNR

Video Name	Block Size	POP			2-D DCT		Gain[1]
		M	Clu.	PCA	M	APSNR (dB)	
			APSNR (dB)	APSNR (dB)			
Ballroom	8x8	6	29.32	31.79	6	31.65	0.14
	16x16	26	30.98	**33.61**	26	33.47	**0.14**
	32x32	102	30.64	34.31	105	34.21	0.1
Vassar	8x8	6	31.63	33.16	6	33.10	0.06
	16x16	26	32.54	**34.73**	26	34.51	0.22
	32x32	102	32.29	35.31	105	35.06	0.25
Exit	8x8	6	32.71	34.98	6	34.93	0.05
	16x16	26	34.22	**37.20**	26	36.34	**0.86**
	32x32	102	34.44	37.89	105	37.45	0.44
BMX	8x8	6	27.07	29.12	6	28.99	0.13
	20x20	40	28.67	**30.53**	40	29.64	**0.89**
	40x40	160	28.82	31.33	167	29.93	1.4

[1]: Gain is between POP with PCA and 2-D DCT

V. CONCLUSION

2-D DCT is playing an important role in image and video compression because it can achieve high compression ratio using universal DCT transform matrix. This paper proposes an alternative approach of POP which adopts variable and adaptive transform matrices. POP's low computational complexity is an attractive advantage over 2-D DCT so that it is beneficial to video applications with high demand on low complexity. Furthermore, the experimental results illustrate that, when PCA is selected as the training tool, the reconstruction quality of the POP-based scheme is comparable to or even better than that of the 2-D DCT based scheme at the same compression ratio. Therefore, POP is promising to assist complex video applications to be implemented on low-cost devices with limited processing capabilities. On the other hand, the POP transform is performed on the source image blocks in this paper; but 2-D DCT can be used for both the source image blocks (intra-frame) and the perdition error blocks (inter-frame) in conventional codecs. So, POP transform is proposed to be used for I-frame coding in conventional codecs or independent key frame coding in distributed video coding [11]. Whether or not POP transform can be applied to perdition error blocks is future work.

REFERENCES

[1] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: applications to image and text data," Proceedings on 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 245–250, 2001.

[2] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform," IEEE Trans. Computers, pp. 90-93, 1974.

[3] H. Anton and C. Rorres, "Elementary Linear Algebra (9th ed.)," John Wiley and Sons, Inc., ISBN 978-0-471-66959-3, 2005.

[4] Y. S. Ho, C. Lee and K. J. Oh, "Overview of view synthesis prediction for multi-view video coding," proceedings of ITC-CSCC 2007, Busan, Korea.

[5] A. Goyal, R. Ren and J. M. Jose, "Feature Subspace Selection for Efficient Video Retrieval," proceedings of the 16th International Conference on Multimedia Modeling, vol. 5916, pp. 725–730, 2010.

[6] P. N. Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining (First Edition)," Addison-Wesley Longman Publishing Inc, Boston, MA, 2005.

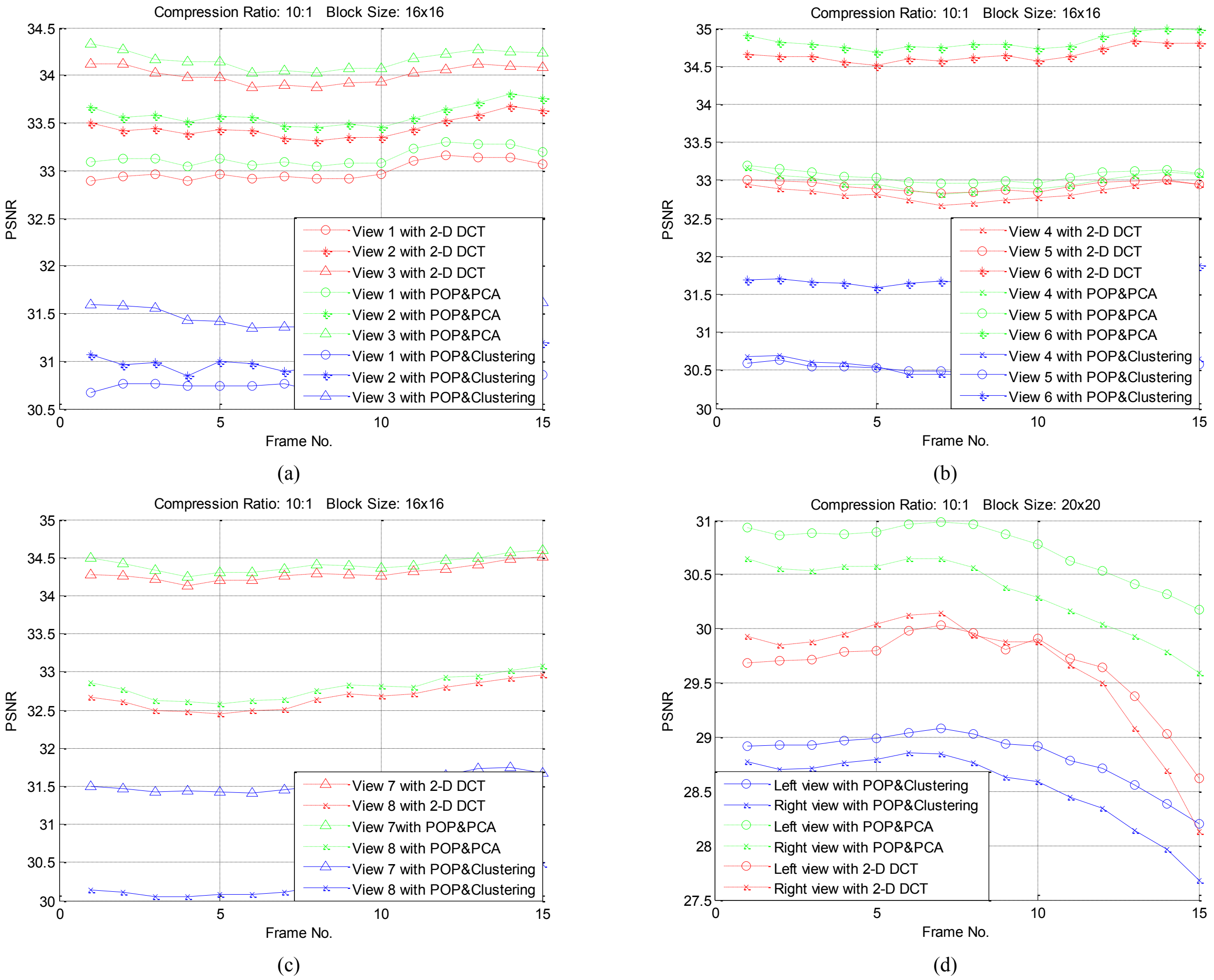
[7] D. Arthur and S. Vassilvitskii, "kmeans++: The advantages of careful seeding," in Proc. ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035, 2007.

[8] C. Yeo and K. Ramchandran, "Robust distributed multi-view video compression for wireless camera networks," IEEE Transactions on Image Processing, vol. 19, pp. 995-1008, Apr. 2010.

[9] Q. Du and J. E. Fowler, "Low-Complexity Principal Component Analysis for Hyperspectral Image Compression," International Journal of High Performance Computing Applications, vol. 22, no. 4, pp. 438-448, doi: 10.1177/1094342007088380, Nov. 2008.

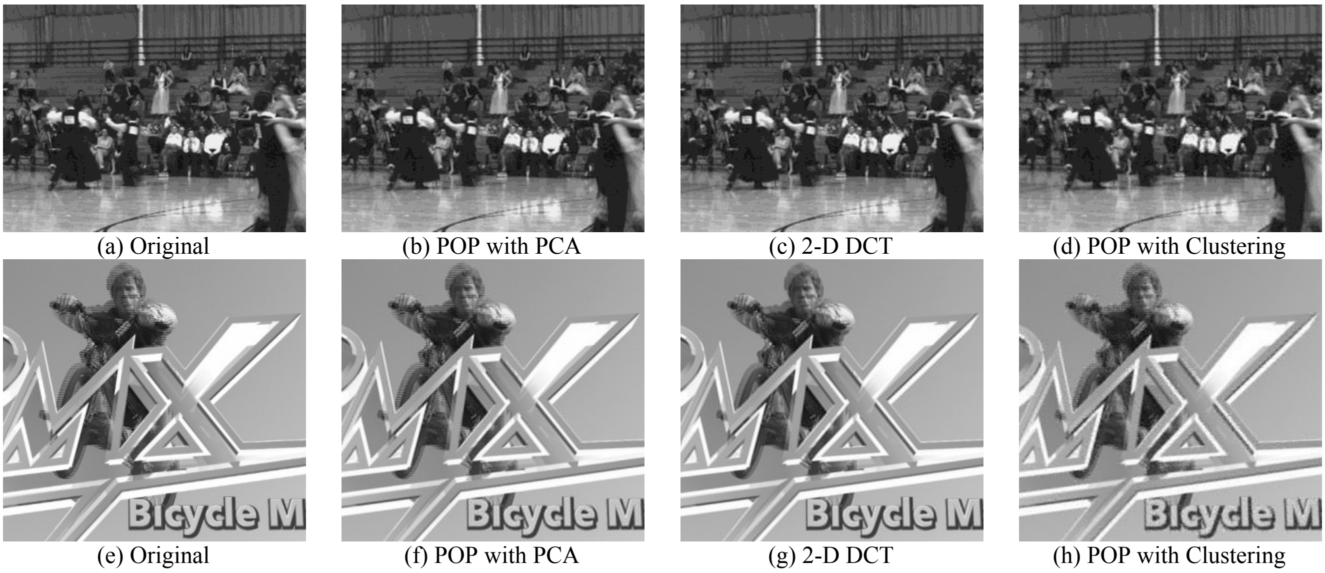
[10] I. F. Akyildiz, T. Melodia and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," Computer Networks, vol. 51, pp. 921-960, Mar. 14, 2007.

[11] A. B. B. Adikari, W. A. C. Fernando and W. A. R. J. Weerakkody, "Independent Key Frame Coding Using Correlated Pixels In Distributed Video Coding," IEE Electronics Letters, vol. 43, issue 7, pp. 387–388, Mar. 29, 2007.



(a) (b) (c) (d)

Figure 3. Frame-by-frame PSNR comparison: Ballroom (a-c) and BMX (d)



(a) Original (b) POP with PCA (c) 2-D DCT (d) POP with Clustering

(e) Original (f) POP with PCA (g) 2-D DCT (h) POP with Clustering

Figure 4. Visual quality comparison: Frame 8 of View 5 of Ballroom (a-d) and cropped Frame 12 of left view of BMX (e-h)
(You can zoom in the images to see more details.)