

This article is downloaded from



CRO CSU Research Output
Showcasing CSU Research

<http://researchoutput.csu.edu.au>

It is the paper published as:

Author: G. Zhouye, W. Lin, B.-s. Lee, C. Tong Lau and M. Paul

Title: Two dimensional singular value decomposition (2D-SVD) based video coding

Year: 2010

Editor: P. W.-C. Siu

Conference Name: IEEE International Conference on Image Processing (ICIP)

Conference Location: USA

Publisher: Institute of Electrical and Electronics Engineers (IEEE) Press

Date: 26-29 September 2010

Abstract: In this paper, we propose a low-complexity video codec based on two-dimensional Singular Value Decomposition (2D-SVD). We exploit the common temporal characteristics of video without resorting to motion estimation. It has been demonstrated that this codec has higher coding efficiency than the relevant existing low complexity codecs. Moreover, the proposed codec performs well to deal with packet loss that is unavoidable in error-prone transmission. Therefore it is with advantages and good potential for wireless video applications such as mobile video calls and wireless surveillance.

DOI/URL: <http://dx.doi.org/10.1109/ICIP.2010.5650998>

http://researchoutput.csu.edu.au/R/-?func=dbin-jump-full&object_id=23103&local_base=GEN01-CSU01

Author Address: mpaul@csu.edu.au

CRO identification number: 23103

TWO DIMENSIONAL SINGULAR VALUE DECOMPOSITION (2D-SVD) BASED VIDEO CODING

Zhouye Gu, Weisi Lin, Bu-sung Lee, Chiew Tong Lau, Manoranjan Paul

School of Computer Engineering, Nanyang Technological University, Singapore, 639798

ABSTRACT

In this paper, we propose a low-complexity video codec based on two-dimensional Singular Value Decomposition (2D-SVD). We exploit the common temporal characteristics of video without resorting to motion estimation. It has been demonstrated that this codec has higher coding efficiency than the relevant existing low complexity codecs. Moreover, the proposed codec performs well to deal with packet loss that is unavoidable in error-prone transmission. Therefore it is with advantages and good potential for wireless video applications such as mobile video calls and wireless surveillance.

Index Terms—Video coding, two-dimensional singular value decomposition, low-complexity, mobile video transmission

1. INTRODUCTION

Hybrid video coding combines the inter-frame prediction and the transformation coding of the prediction residues. Many international video coding standards, such as H.263, H.264/AVC, MPEG-1/2/4[1], adopt this coding framework. Temporal redundancy is exploited by the inter-frame prediction techniques (e.g. motion estimation). As a result, compression efficiency is significantly improved.

However, the cost of this high compression efficiency is the high complexity due to the motion estimation process which leads to the high battery/power consumption, low speed of software implementation and high cost of hardware implementation [2]-[4]. All these make the use of the hybrid codecs in mobile devices and communications difficult.

Motion JPEG and Motion J2K [5] are the standard video codecs with low computational complexity. However, they do not exploit temporal redundancy in video to achieve higher compression. Alternative schemes have been explored without motion estimation for video coding. Chiu and Berger proposed to use differential frame replenishment for videoconferencing [13]. Law and Nguyen proposed a low-complexity video codec named Motion Wavelet Difference Reduction (MWDR) codec [4] for the

application of mobile video calls. By coding the difference between two consecutive frames with Wavelet difference reduction, their codec achieves much higher coding efficiency than Motion JPEG [4].

The 1D-SVD (*Singular Value Decomposition*) based coding has also been proposed due to the desirable attribute of SVD: optimized energy compaction [6]. With SVD transformation, the coefficient matrix, containing the square root of eigenvalues, is diagonalized. This coefficient matrix contains much less non-zero coefficients compared with the coefficient matrices of other transformations. However, the 1D-SVD based coding techniques typically achieve only modest compression because the eigenvectors must be coded along with the associated eigenvalues [7,8].

To make use of the energy compaction property of SVD based coding while overcoming the inefficiency of coding the eigenvectors, in this paper, we propose a new low-complexity codec based on two-dimensional Singular Value Decomposition (2D-SVD) [11,12]. Compared with the 1D-SVD codecs [7,8], the proposed 2D-SVD codec inherits the energy compaction property from a 1D-SVD scheme, while it needs to code and transmit much fewer coefficients. In Section 2 of this paper, the 2D-SVD and its characteristics relevant to our work are to be firstly introduced. The proposed 2D-SVD codec is then presented as Section 3. Based on its optimal energy compaction property, the 2D-SVD codec outperforms both Motion J2K and MWDR codecs in terms of coding video quality. This has been confirmed with experiments in Section 4. Our codec also shows its robustness under different packet loss rates. The last section concludes this paper.

2. OVERVIEW OF 2D-SVD

2.1. Principles of 2D-SVD

The low-rank approximation of matrices has recently received much attention for research efforts [9]-[12]. The 2D-SVD solution belongs to the category of Simultaneous Low Rank Approximation of Matrices (SLRAM) [12]:

$$\min_{U_l, U_r, M_i} \sum_{i=1}^n \|A_i - U_l M_i U_r^T\|_F^2 \quad (1)$$
$$s. t. \quad U_l^T \cdot U_l = I_n, U_r^T \cdot U_r = I_2$$

This work was supported by MoE AcRF Tire 2, Singapore, Grant Number: T208B1218.

where $A_i \in \mathbb{R}^{r \times c}$ is the i -th frame/image block of video and we aim to compute two matrices $U_l \in \mathbb{R}^{r \times l_1}$, $U_r \in \mathbb{R}^{c \times l_2}$, such that $I_{l_1} \in \mathbb{R}^{l_1 \times l_1}$, $I_{l_2} \in \mathbb{R}^{l_2 \times l_2}$, $M_i \in \mathbb{R}^{l_1 \times l_2}$, $l_1 \leq r$ and $l_2 \leq c$

The problem in (1) is equivalent to

$$\max_{U_l, U_r} \sum_{i=1}^n \left\| U_l^T A_i U_r \right\|_F^2 \quad (2)$$

$$s. t. \quad U_l^T \cdot U_l = I_{l_1}, U_r^T \cdot U_r = I_{l_2}$$

A near-optimal solution for such a SLRAM problem (2) is given in [11] and our 2D-SVD codec to be proposed in details in Section 3 is based on this approach.

The 2D-SVD procedures are therefore described in the following steps [11]:

Given a GOP (Group of Pictures) of n frames, we denote the i -th frame as A_i . The mean frame, A_{mean} , is abstracted from each frame and consequently we have $\hat{A}_i = A_i - A_{mean}$.

The row-row and column-column covariance matrices are defined as $F = \sum_{i=1}^n \hat{A}_i \hat{A}_i^T$, and $G = \sum_{i=1}^n \hat{A}_i^T \hat{A}_i$. Both U_l and U_r are made up by k principle eigenvectors of F and s principle eigenvectors of G , respectively, i.e.,

$$F = \sum_{p=1}^r \lambda_p u_p u_p^T \quad U_l \equiv (u_1, \dots, u_k)$$

$$G = \sum_{q=1}^c \zeta_q v_q v_q^T \quad U_r \equiv (v_1, \dots, v_s)$$

Let $M_i = U_l^T \hat{A}_i U_r$, and $\bar{A}_i = U_l M_i U_r^T$. We calculate the near-optimal approximation of each block A_i as $\hat{A}_i = \bar{A}_i + A_{mean}$

According to the experiment in [12], the minimum squared error is achieved at $s/k \approx 1$. Thus, we set $s = k$ in our analysis.

2.2. Comparison of 2D-SVD and 1D-SVD

As we showed above, for a group of frames/image blocks $A_1 \sim A_n$, 2D-SVD coding only needs to transmit two unitary matrices U_l, U_r and a group of coefficient matrices $M_1 \sim M_n$. On the other hand, for 1D-SVD coding each frame/image block A_i , we need to transmit two unitary matrices U_l, U_r and a coefficient matrix M_i . We illustrate 1D-SVD and 2D-SVD in Figure 1.

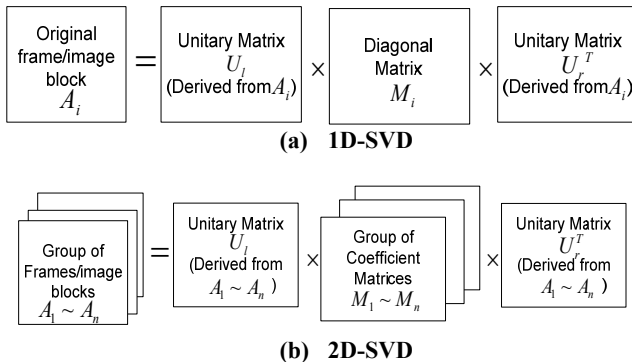


Figure 1. Comparison of 2D-SVD and 1D-SVD

Table 1. Number of Coefficients for 1D-SVD and 2D-SVD

Block Size	Method	Video Resolution			
		QCIF	CIF	2CIF	4CIF
8 × 8	1D-SVD	1.34 × 10 ⁶	5.38 × 10 ⁶	1.07 × 10 ⁷	2.15 × 10 ⁷
	2D-SVD	6.84 × 10 ⁵	2.73 × 10 ⁶	5.47 × 10 ⁶	1.09 × 10 ⁷
Coefficients saved		49%			
16 × 16	1D-SVD	1.30 × 10 ⁶	5.23 × 10 ⁶	1.04 × 10 ⁷	2.09 × 10 ⁷
	2D-SVD	6.84 × 10 ⁵	2.73 × 10 ⁶	5.47 × 10 ⁶	1.09 × 10 ⁷
Coefficients saved		48%			

Each frame of video can be divided into several blocks and then each block is coded with a 1D-SVD codec [6]-[8]. Suppose we want to compress a video with a GOP of n frames with the frame resolution of $M \times N$, and the block size is $m \times m$. For 1D-SVD, $nMN(2m^2 + m)/m^2$ coefficients are needed, while for 2D-SVD, $(2+m)MN$ coefficients are required. In Table 1 with the GOP size being 24, we can observe that at various resolutions and block sizes 2D-SVD needs only about 50% of coefficients compared with 1D-SVD.

In a 1D-SVD scheme, U_l and U_r need to be transmitted apart from M_i , for an image block of every frame. In the 2D-SVD scheme, the common features of the image block along the time axis are represented by U_l and U_r , which are needed to be transmitted only once for a segment of video; the changing features with time are captured by M_i which is largely diagonalized. This is the motivation and the reason of improvement for using 2D-SVD in this work.

2.3. Computational complexity comparison of hybrid video codec and 2D-SVD video codec

The main difference of computational complexity between the 2D-SVD coding and hybrid coding is the 2D-SVD decomposition and motion estimation (ME). Still suppose we have a GOP of n frames and the block size is $m \times m$. For a $m \times m$ matrix, it takes about $13m^3$ operations to calculate the two eigenvector matrices [12],[14]. Since we need $m^2(2n+1)$ operations to calculate and subtract the mean frame, the overall operations required for 2D-SVD codec is $13m^3 + m^2(2n+1)$. For full-search ME, given a search window $W \times W$, the operations required are $2nW^2m^2$ [4]. If the GOP size is 24 and the block size and search window size are 16×16 and 32×32 respectively, by our estimation, ME takes 191 times more operations than that of 2D-SVD decomposition. This enables the proposed method to be adopted by devices which can not afford the intensive computation of ME but with higher coding efficiency than the existing non-ME based codecs.

3. FRAMEWORK OF 2D-SVD CODEC

Figure 2 depicts the system diagram of the proposed 2D-SVD encoder, with each part discussed below.

Once we abstract the mean frame, we compress the mean frame with JPEG standard by quality factor of 95.

After mean frame subtraction, a frame is normalized so that after normalization, the energy (the Frobenius norm, F-Norm) of A'_i is 1. The F-Norm of each frame, $\|A'_i\|_F$, is transmitted within the headerfile of the corresponding compressed frame with 6 Bytes.

Then, we divide the normalized frames into 16×16 macro-block groups (MBGs). For the j -th MBG B_j , we conduct 2D-SVD on the resultant MBG and we get the corresponding eigenvector matrices U_l^j, U_r^j and the group of 16×16 coefficient matrices $M_1^j \sim M_n^j$. For the eigenvector matrices, we just quantize all scale values of each eigenvector with 8 bits. The mean frame and the eigenvector matrices are compacted in the Group Information (GI) file. We quantize the group coefficient matrices using 16×16 quantization table defined in (3) below, accounting for the fact that the coefficients located around the top left corner contain most energy, while those at the right bottom part contain much less energy.

Let $C(i, j)$ be the entry of the quantization table at location (i, j) , where $1 \leq i \leq 16, 1 \leq j \leq 16$. Then,

$$C(i, j) = 0.001 \times \begin{cases} 3.0 & \text{if } 2 \leq i + j \leq 9 \\ 4.5 & \text{if } 9 < i + j \leq 17 \\ 6.0 & \text{else} \end{cases} \quad (3)$$

Finally, by entropy coding, the coefficient matrix of each frame is further compressed and we transmit the coefficient matrices frame by frame.

The decoder part simply carries out the reverse steps as the encoder. Firstly, the GI is decoded and the mean frame and eigenvector matrices are extracted. Then we get the reconstructed frames without the mean frame by the inverse-SVD and multiplying the resultant frame with its corresponding frame energy $\|A'_i\|_F$. Finally, by adding the reconstructed mean frame, we obtain the reconstructed frame A_i .

4. EXPERIMENT RESULTS

4.1. Coding efficiency comparison of low complexity video codecs

We compare the coding efficiency of the proposed 2D-SVD codec, the MWDR codec [4] and the Motion J2K codec in

Table 2. PSNR (dB) comparison of three low complexity codecs (The results for MWDR codec are extracted from [4]; the numbers in bold indicate the best cases among the three codecs under the same condition)

Bit Rate	Video	2D-SVD	MWDR	Motion J2K
250 Kbps	Grandma	38.9	37.6	33.3
	Salesman	37.2	36.7	29.5
	Claire	40.3	37.7	37.3
150Kbps	Grandma	36.8	36.5	31.1
	Salesman	35.6	35.9	27.9
	Claire	38.4	37.4	34.2
Average PSNR improvement of the proposed codec			0.90	5.65

Table 2. For the proposed 2D-SVD codec, the 2D-SVD is carried out on the GOP containing 24 frames and the frame rate is 25 fps (the same configuration as [4]).

Table 2 shows that the proposed 2D-SVD codec outperforms the Motion J2K codec by 5.65 dB on average. The proposed 2D-SVD codec also outperforms the MWDR codec except for the case of ‘‘Salesman’’ at 150Kbps, in which the latter is slightly better; the average PSNR improvement is 0.9dB over the MWDR codec. The improvement is higher with Motion J2K since the Motion J2K codec does not explore temporal redundancy.

4.2. Performance comparison with packet loss

Since the MWDR codec codes the difference of two frames, in the transmission through error-prone channel, if a frame is lost and replaced by the previous frame, the consequent frames will have the incorrect reference frame, and thus the error is accumulated and amplified as the decoding process goes on. Therefore, we can expect that the MWDR codec or other codecs built with similar principle will suffer more quality degradation than the 2D-SVD codec in the packet loss test. For the 2D-SVD codec, a frame only depends on the GI but not other frames so we can expect it more robust in error-prone transmission. Since Motion J2K codec is known to have better error resilience due to its frame independency with decoding, we have conducted the comparison between the proposed 2D-SVD coder with the Motion J2K in Table 3. We show the error resilience property of the 2D-SVD codec in the packet loss test.

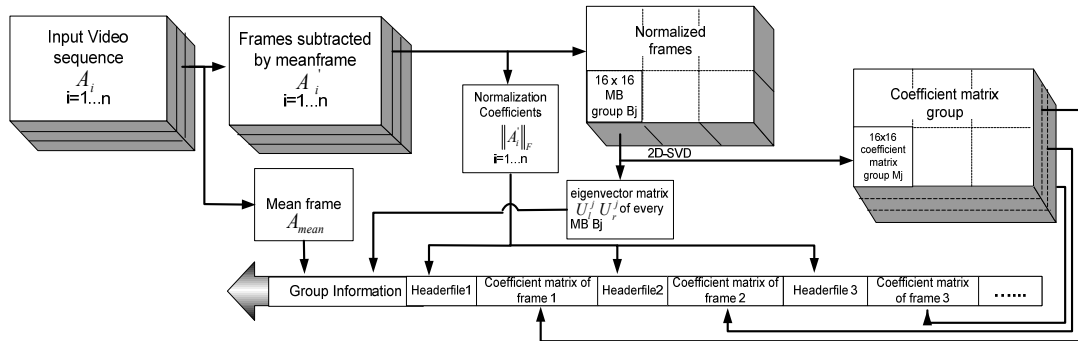


Figure 2. The proposed 2D-SVD Encoder

Table 3. PSNR (dB) Comparison in Packet Loss Test
(The numbers in bold indicate the best cases between the two codecs under the same condition)

Consequence name	Bit Rate	Video codec	Packet Loss Rate				
			0%	1%	3%	5%	10%
Hall	150Kbps	Motion J2K	26.1	25.9	25.6	25.4	25.1
		2D-SVD	30.2	30.0	29.2	28.2	27.1
	250Kbps	Motion J2K	29.3	29.2	28.8	28.5	28.1
		2D-SVD	33.7	33.5	32.9	32.1	30.8
Grandma	150Kbps	Motion J2K	31.1	31.0	30.7	30.3	29.7
		2D-SVD	36.7	36.5	35.7	34.9	33.4
	250Kbps	Motion J2K	33.3	33.2	32.9	32.6	32.1
		2D-SVD	38.6	38.5	37.6	36.5	34.9
Salesman	150Kbps	Motion J2K	27.9	27.9	27.7	27.5	27.0
		2D-SVD	35.5	35.2	34.3	33.5	32.0
	250Kbps	Motion J2K	29.5	29.4	29.2	28.9	28.6
		2D-SVD	37.0	36.9	35.5	34.2	32.9
Claire	150Kbps	Motion J2K	34.2	34.2	34.0	33.7	33.3
		2D-SVD	38.0	37.8	37.0	36.2	34.9
	250Kbps	Motion J2K	37.3	37.2	36.9	36.6	36.2
		2D-SVD	40.1	39.9	38.9	37.8	36.4

If a frame packet is lost, we use the nearest decoded frame to conceal that lost frame for both 2D-SVD and Motion J2K. The GI file has been interleaved and a lost GI block is reconstructed by averaging its neighboring blocks in the 2D-SVD coder. The IP packet header is assumed to be 20Bytes and is considered in the bit rate calculation.

In Table 3, as expected, Motion J2K codec shows its advantages of inter-frame independency. For these low motion videos, when packet loss rate increases from 0% to 10%, the PSNR of Motion J2K codec only decreases about 1dB on average. For the 2D-SVD codec, the decrease of PSNR is higher, since it is conditionally inter-frame independent. However, even at 10% packet loss rate, the overall PSNR of the 2D-SVD is still higher than that of Motion J2K for the same bit rate, due to the substantial gain in the 2D-SVD coding itself.

5. CONCLUSION

In this paper, we have explored the 2D-SVD (*two-dimensional Singular Value Decomposition*) for low-complexity video coding, without adopting motion estimation. The proposed codec has higher coding efficiency than other low-complexity video codecs due to the good energy compaction property of SVD. In comparison with the Motion J2K and related other codecs, which also do not adopt motion estimation, the proposed codec outperforms significantly in coding picture quality at the same bit rate. Even with packet loss, the overall performance of the proposed codec is better than the Motion J2K which is with total inter-frame independency. The proposed codec is therefore suitable for scenarios of mobile video calls and wireless surveillance where low-complexity and good error resilience are required.

6. REFERENCES

- [1] J. Watkinson, *The MPEG handbook: MPEG-1, MPEG-2, MPEG-4*, Focal Press, Boston, 2001.
- [2] A. Bahari, T. Arslan, A. Erdogan, "Low-Power H.264 Video Compression Architectures for Mobile Communication," *IEEE Trans. CSVT.*, vol. 19(9), pp. 1251–1261, 2009.
- [3] C. Chen, S. Chien, Y. Huang, T. Chen, T. Wang, L. Chen, "Analysis and Architecture Design of Variable Block-Size Motion Estimation for H.264/AVC," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol.53(3), pp. 578–593, 2006.
- [4] Y. L. Law and T. Q. Nguyen, "Motion wavelet difference reduction (MWDR) video codec," *IEEE International Conference on Image Processing*, vol 4, pp. 2303 - 2306 2004.
- [5] Information Technology – JPEG 2000 Image Coding System, Part 3: Motion JPEG 2000, ISO/IEC 15444-3:2002.
- [6] H. Andrews and C. Patterson, "Singular value decomposition (SVD) image coding," *IEEE Trans. Commun.*, pp. 425-432, 1976.
- [7] H. Ochoa, K.R. Rao, "A Hybrid DWT-SVD Image Coding System (HDWTSVD) for Monochromatic Images," *SPIE's 15th Annual Symposium*, 2003.
- [8] T. Saito and T. Komatsu, "Improvement on Singular Value Decomposition Vector Quantization," *Electronics and Communications in Japan*, Part 1, vol. 73, pp. 11-20, 1990.
- [9] C. Ding, H. Huang, and D. Luo, "Tensor reduction error analysis – applications to video compression and classification," *IEEE Conf. on CVPR*, pp. 1-8, 2008.
- [10] K. Inoue and K. Urahama. "Equivalence of non-iterative algorithms for simultaneous low rank approximations of matrices," *IEEE Conf. on CVPR*, pp. 154-159, 2006.
- [11] C. Ding and J. Ye. "Two-dimensional singular value decomposition (2dsvd) for 2d maps and images," *Int'l Conf. Data Mining*, pp. 32–43, 2005
- [12] J. Ye. "Generalized low rank approximations of matrices," *Machine Learning*, vol. 61, pp. 167-191, 2005.
- [13] Y. Chiu, T. Berger. "A software-only video codec using pixelwise conditional differential replenishment and perceptual enhancements," *IEEE Trans. CSVT*, Vol.9(3), pp. 438–450, 1999.
- [14] www.navab.cs.tum.edu - *3D Computer Vision Script Draft*