QUALITY ENHANCEMENT FOR MOTION JPEG USING TEMPORAL REDUNDANCIES

Dũng T. Võ and Truong Q. Nguyen

ECE Department, UCSD, La Jolla CA 92093-0407
http://videoprocessing.ucsd.edu
Email: d3vo@ucsd.edu, nguyent@ece.ucsd.edu

ABSTRACT

We propose a post processing algorithm to enhance the quality of Motion JPEG (MJPEG) by exploiting temporal redundancies. The error between the estimated and original blocks is analyzed using the translational relation in the discrete cosine transform (DCT) domain. The proposed algorithm permits reconstructing the high frequency coefficients lost during quantization and therefore reduces the ringing artifact. Blocking artifact reduction is verified by the decrease in the variance of this coefficient error. Results in quality and PSNR improvement for both cases of integer and sub-pixel motion vectors are verified by simulations on video sequences.

***Index Terms*—** Discrete cosine transform (DCT), Motion JPEG (MJPEG), blocking artifact, ringing artifact.

1. INTRODUCTION

Motion JPEG (MJPEG) compresses separately each frame of the video sequence in JPEG format. Quality enhancement for MJPEG until now has focused on improving the quality of each single JPEG frame. Most quality problems concerning JPEG are blocking and ringing artifacts, especially at low bitrate compression.

One method to deal with blocking artifacts is using lapped orthogonal transform (LOT) [1], which increases the dependence between adjacent blocks. To be compatible with the JPEG standard, many pixel-based and DCT-based post processings methods have been proposed instead of using LOT. [2] applied an adaptive median filter to remove the high frequencies caused by unwanted edges between adjacent blocks. Other pixel-based methods are maximum a posteriori probability approach (MAP) [3], constrained least squares (CLS) and projection on convex sets (POCS) [4], which require many iterations with high computational load. The DCT-based methods [5] [6] adjusted the quantized DCT coefficients to reduce quantization error. To decrease ringing artifacts, [7] and [8] used different operators for edge detection and applied a suitable filter for pixels near edges. These methods did not solve the problem completely because the high frequency components of the resulted images were not reconstructed.

Due to the strong correlation between adjacent frames in video sequence, information from successive frames can be used to reduce the quantization error resulted from truncating the DCT coefficients of each frame. This paper proposes a novel method of using the previous and future frames to enhance the quality of the current frame.

This work is supported by Texas Instruments and matching fund from UC Discovery program.

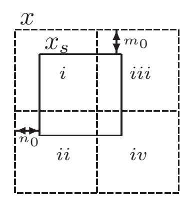


Fig. 1. Translation between blocks of image x_s and x

2. TRANSLATIONAL RELATION IN DCT DOMAIN

In MJPEG, each frame is processed in separate blocks of size $N \times N$ ($N = 8$). Assume one block of frame x_s matches to another block located among 4 blocks of frame x as in Fig. 1. The DCT transform (type II) of x_s can be obtained by

$$X_s(u,v) = \frac{2}{N} k_u k_v \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x(m+m_0, n+n_0) C_u^m C_v^n$$
$$\text{for } m,n,u,v = 0,\ldots,N-1 \quad (1)$$

where $C_w^p = \cos\left(\frac{(2p+1)w\pi}{2N}\right), S_w^p = \sin\left(\frac{(2p+1)w\pi}{2N}\right)$

and $k_j = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } j = 0 \\ 1 & \text{otherwise} \end{cases}$

In (1), if we replace x by its DCT coefficients X, X_s can be calculated based on the DCT coefficients of 4 blocks (i, ii, iii and iv) of image x or more generally, based on a function $F(X, m_0, n_0)$. After DCT transform, the quantization process truncates the DCT coefficients X to obtain the output X_q with error ΔX_q. Because of the non-linear characteristic of the quantization function, the error $\Delta X_q(u,v)$ prevents the quantized DCT coefficients of the original block X_q and the shifted blocks $X_{s,q}$ from satisfying (1). Linearizing the quantization function, a displacement $X^d_{s,q}$ of $X_{s,q}$ can be estimated based on X_q

$$X^d_{s,q} = F\big(X_q, m_0, n_0\big) \tag{2}$$

Instead of having zero values due to truncation, high frequency DCT coefficients are re-created using (2). This equation is equivalent to a simple translation in the pixel domain

$$x^d_{s,q}(m,n) = x_q(m+m_0, n+n_0) \tag{3}$$

3. QUALITY ENHANCEMENT USING TEMPORAL REDUNDANCIES

In Fig. 3, motion estimation (ME) is used to find the motion vector between blocks of frames. Assume that the present block is shifted from one block of the previous and future frames respectively by (m^b_0, n^b_0) and (m^f_0, n^f_0) pixels, the backward estimated version $x^b_q(t,m,n)$ and forward estimated version $x^f_q(t,m,n)$ of compressed block $x_q(t,m,n)$ are calculated by using (3). Consequently, an averaging scheme is used to calculate the final processed block

$$x'_q(t,m,n) = \frac{1}{3}\big(x^f_q(t,m,n) + x_q(t,m,n) + x^b_q(t,m,n)\big) \tag{4}$$

This is equivalent to a temporal filter for aligned blocks of different frames. Because DCT is an orthonormal transform, the mean square error between the estimated and original blocks can be obtained by their DCT coefficients

$$E'_q(t,u,v) = X'_q(t,u,v) - X(t,u,v) \tag{5}$$

Consider Q as the quantization matrix, the variance of E'_q is calculated with the assumption that the quantization errors are white and independent for DCT coefficients in the same block, in different blocks and different frames

$$\sigma^2_{E'_q}(u_0,v_0) = \frac{1}{9}\frac{Q^2(u_0,v_0)}{12} + \frac{1}{9}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\sum_{l=i}^{iv}\frac{Q^2(u,v)}{12} \times\Big(\big(K^b_l(u,v)\big)^2 + \big(K^f_l(u,v)\big)^2\Big) \tag{6}$$

$$\text{where} \quad K^b_{iv}(u,v) = (-1)^{u_0+v_0}\sum_{m=0}^{m^b_0-1}\sum_{n=0}^{n^b_0-1} J^b(m,n) \tag{7}$$

$$K^b_{iii}(u,v) = (-1)^{v_0}\sum_{m=m^b_0}^{N-1}\sum_{n=0}^{n^b_0-1} J^b(m,n) \tag{8}$$

$$K^b_{ii}(u,v) = (-1)^{u_0}\sum_{m=0}^{m^b_0-1}\sum_{n=n^b_0}^{N-1} J^b(m,n) \tag{9}$$

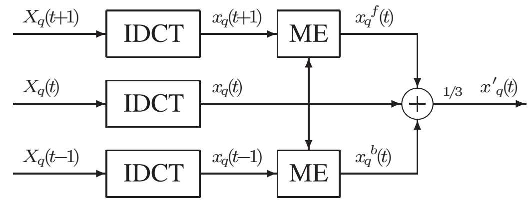


Fig. 2. Block diagram of the enhancement algorithm

$$K^b_i(u,v) = L^b(u,v) + \sum_{m=m^b_0}^{N-1}\sum_{n=n^b_0}^{N-1} J^b(m,n) - \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} J^b(m,n) \tag{10}$$

$$\text{where} \quad J^b(m,n) = \frac{4}{N^2}k_{u_0}k_{v_0}k_u k_v C^m_u C^n_v C^{m-m^b_0}_{u_0} C^{n-n^b_0}_{v_0} \tag{11}$$

$$L^b(u_0,v_0) = C^{m^b_0-\frac{1}{2}}_{u_0} C^{n^b_0-\frac{1}{2}}_{v_0} \tag{12}$$

$$L^b(u\neq u_0, v\neq v_0) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} J^b(m,n)\Big(C^m_{u_0}S^n_{v_0}\frac{C^{m^b_0-\frac{1}{2}}_{u_0} S^{n^b_0-\frac{1}{2}}_{v_0}}{C^{m-m^b_0}_{u_0} S^{n-n^b_0}_{v_0}} + \frac{S^m_{u_0} S^{m^b_0-\frac{1}{2}}_{u_0}}{C^{m-m^b_0}_{u_0} S^{n-n^b_0}_{v_0}}\big(C^n_{v_0} C^{n^b_0-\frac{1}{2}}_{v_0} + S^n_{v_0} S^{n^b_0-\frac{1}{2}}_{v_0}\big)\Big) \tag{13}$$

and similiarly for $K^f_l(u,v)$.

To get improvement, $\sigma^2_{E'_q}(u_0,v_0)$ must be less than the variance of the original error $\frac{Q^2(u_0,v_0)}{12}$. This is equivalent to

$$\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\sum_{l=i}^{iv}\Big(\big(K^b_l(u,v)\big)^2 + \big(K^f_l(u,v)\big)^2\Big)\frac{Q^2(u,v)}{Q^2(u_0,v_0)} \leq 8 \tag{14}$$

In real video sequence, pure translation is rarely satisfied. Assume that the block in x_s is the shifted version of one block of image x with difference Δx

$$x_s(m,n) = x(m+m_0, n+n_0) + \Delta x(m+m_0, n+n_0) \tag{15}$$

With the assumption that Δx is zero mean noise with variance $\sigma^2_{\Delta x}$ and is independent to x

$$\sigma^2_{E'_{q,real}}(u_0,v_0) = \sigma^2_{E'_q}(u_0,v_0) + k^2_{u_0}k^2_{v_0}\sigma^2_{\Delta x} \tag{16}$$

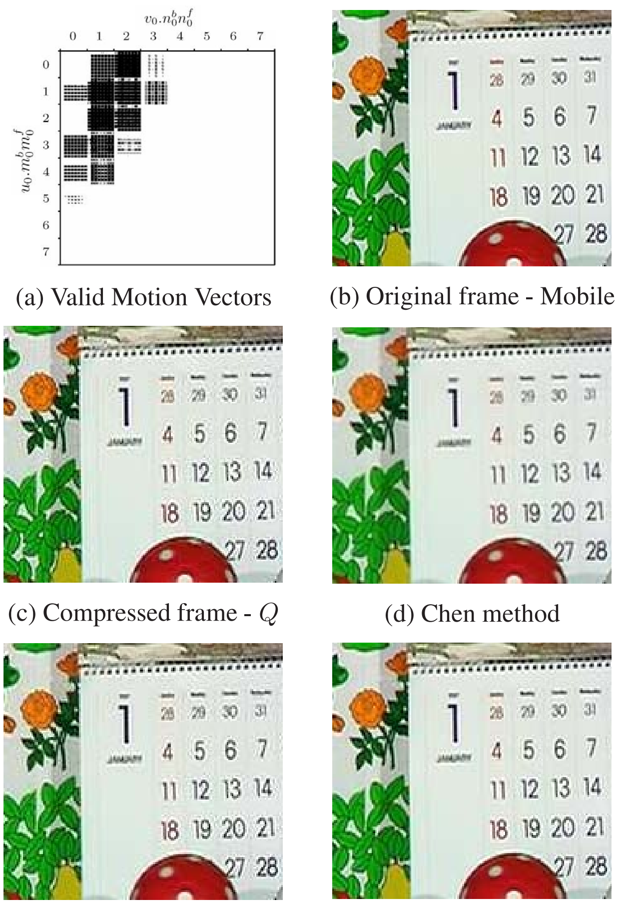
The error variance in this case increases by $k^2_{u_0}k^2_{v_0}\sigma^2_{\Delta x}$ comparing with the error variance of the ideal translation. The non-ideal relation also decreases the accuracy in detecting the motion vectors between blocks.

A more general scheme which uses an arbitrary number of previous and future frames to enhance the current frame can be considered using the same process. The final processed block is obtained by applying an adaptive filter to the estimated versions. The optimal coefficients of the filter are calculated by minimizing the variance of the error E'_q under the same assumption.

4. SIMULATION RESULTS

4.1. Enhancement in pure translational video sequence

Theoretically, only those motion vectors which satisfy (14) will provide enhancement in image quality. Fig. 3(a) shows all cases of motion vectors from $(0, 0)$ to $(7, 7)$ with block size 8×8. Both coordinates are in base eight, such as on the horizontal axis, $x = 2.43$ means that $u_0 = 2, m_0^b = 4$ and $m_0^f = 3$ and similiarly for vertical axis. If the point at $(u_0.m_0^b m_0^f, v_0.n_0^b n_0^f)$ meets the condition, its color is white and otherwise. At high frequency, all motion vectors give improvement in PSNR, but at middle frequency, only a few motion vectors provide enhancement.



(a) Valid Motion Vectors (b) Original frame - Mobile

(c) Compressed frame - Q (d) Chen method

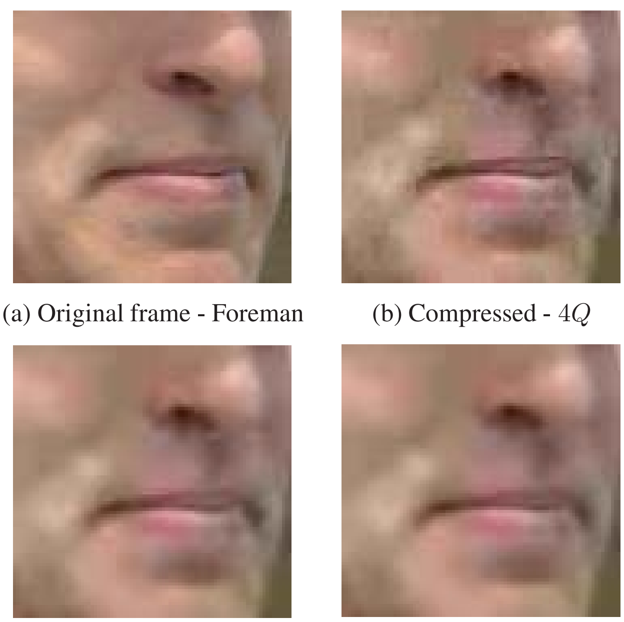
(e) Liu method (f) Proposed method

Fig. 3. Enhancement for ideal case - Low compression.

The simulation for ideal case is applied to a single frame of different video sequences. Previous and future frames are translated from the present frame respectively by $(1, 1)$ and $(-1, -1)$ pixel. Motion vectors are found using Full Search algorithm with window size 9×9. Quantization error prevents detecting correctly all motion vectors. The minimum block distortion has to be less than a threshold $= 5000$ to be a valid motion vector for pure translation. Relation in (2) permits replacing high frequency DCT coefficients by a better replacement with less error variance and helps reducing ringing artifact. Comparing with the enhanced frames in Fig. 3(d) and (e) using Chen and Liu methods, Fig. 3(f) has higher PSNR and reduces more ringing artifacts near sharp edges. Table 1 validates the effectiveness of the proposed method for different sequences in case of using quantization matrix Q.

Table 1. PSNR enhancement in dB for ideal sequences

Sequence	Decoded	Chen	Liu	Proposed
Foreman	32.72	32.82	32.63	**33.50**
Mobile	25.69	25.39	25.51	**26.96**
City	31.17	31.01	30.83	**32.81**
Bus	30.16	29.91	29.61	**31.92**
Lena	35.83	35.82	35.55	**37.31**



(a) Original frame - Foreman (b) Compressed - $4Q$

(c) Chen method (d) Proposed combined method

Fig. 4. Enhancement ideal case - High compression.

With higher level compression ($4Q$), the compressed frame is more affected by blocking artifact as in Fig. 4(b) ($28.08dB$). This artifact is reduced in Fig. 4(c) ($28.60dB$) with Chen method. Due to high quantization error, the proposed method is applied to image resulted from Chen method to get more accurate motion vectors. The resulting image in Fig. 4(d) ($28.85dB$) has less ringing and blocking artifacts.

4.2. Enhancement in real video sequence

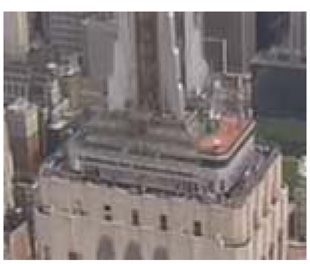
The proposed algorithm is applied to the City sequence, the motion vectors are found to $1/8$ pixel accuracy. Sub-pel block matching requires interpolation for the whole frame, but the enhanced method is only need implementing at integer pixels. The enhanced frames in Fig. 5(c), (d) and (e) have less ringing artifacts than the JPEG image in Fig. 5(b). The graph in Fig. 5(f) verifies the robustness of the algorithm for the whole sequence. The results in Table 2 show the enhancement for both cases of quantization matrix Q and $2Q$. There is only a small enhancement in PSNR between using $1/2$, $1/4$ and $1/8$ pixel accuracy, so half pixel ME is sufficient for many applications. All full size images and video results can be found at http://videoprocessing.ucsd.edu/∼dungvo/ICIP07/

(a) Original frame - Mobile

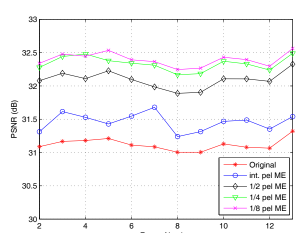
(b) Compressed frame

(c) Enhanced - integer pel ME

(d) Enhanced - $\frac{1}{2}$ pel ME



(e) Enhanced - $\frac{1}{4}$ pel ME



(f) PSNR for City frames

Fig. 5. Enhancement for real case.

5. CONCLUSION

A post processing algorithm to improve the quality for Motion JPEG is proposed. The algorithm reduces the ringing and blocking artifacts and is verified in both PSNR and visual quality. Since the proposed method uses interframe correlation, it can be combined with other spatial-based methods, especially in high level compression, where ME is not accurate. In practice, this method only requires ME that is available in MPEG encoders. Future work will extend the results using a more general motion model between blocks and other video coding standards.

Table 2. PSNR enhancement for 3rd frame of City sequence

Quant.	JPEG	int. pel	$\frac{1}{2}$ pel	$\frac{1}{4}$ pel	$\frac{1}{8}$ pel
Q	31.18	31.62	32.19	32.44	32.48
2Q	28.79	29.47	29.87	29.98	30.00

6. REFERENCES

[1] H.S. Malvar and D.H. Staelin, "The LOT: Transform coding without blocking effects," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 553–559, April 1989.

[2] B. Ramamurthi and A. Gersho, "Nonlinear space-variant postprocessing of block coded images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 1258–1268, October. 1986.

[3] R.L. Stevenson, "Reduction of coding artifacts in transform image coding," *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol. 3, pp. 401–404, 1993.

[4] Y. Yang, N.P. Galatsanos, and A.K. Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 421–432, December. 1993.

[5] T. Chen, H.R. Wu, and B. Qiu, "Adaptive postfiltering of transform coefficients for the reduction of blocking artifacts," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 594–602, May. 2001.

[6] S. Liu and A.C. Bovik, "Efficient DCT-domain blind measurement and reduction of blocking artifacts," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 1139–1149, December. 2002.

[7] H.S. Kong, A. Vetro, and S. Huifang, "Edge map guided adaptive post-filter for blocking and ringing artifacts removal," *Proc. IEEE Int. Sym. Circuits and Syst.*, vol. 3, pp. 929–932, 2004.

[8] S.H. Oguz, Y.H. Hu, and T.Q. Nguyen, "Image coding ringing artifact reduction using morphological postfiltering," *Proc. IEEE Int. Work. Multimedia Signal Processing*, pp. 628–633, 1998.