NEW MOTION COMPENSATION MODEL VIA FREQUENCY CLASSIFICATION FOR FAST VIDEO SUPER-RESOLUTION

Kwok-Wai Hung and *Wan-Chi Siu*

Centre for Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University

ABSTRACT

A typical dynamic reconstruction-based super-resolution video involves three independent processes: registration, fusion and restoration. Fast video super-resolution systems apply translational motion compensation model for registration with low computational cost. Traditional motion compensation model assumes that the whole spectrum of pixels is consistent between frames. In reality, the low frequency component of pixels often varies significantly. We propose a translational motion compensation model via frequency classification for video super-resolution systems. A novel idea to implement motion compensation by combining the up-sampled current frame and the high frequency part of the previous frame through the SAD framework is presented. Experimental results show that the new motion compensation model via frequency classification has an advantage of 2dB gain on average over that of the traditional motion compensation model. The SR quality has 0.25dB gain on average after the fusion process which is to minimize error by making use of the new motion compensated frame.

Index Terms — Super-resolution, Dynamic video SR, new motion compensation, fusion

1. INTRODUCTION

Image super-resolution (SR) is well studied in the literature. Typical blind and non-blind image super-resolution algorithms can provide good quality results. In contrast, good quality always means high computational cost [1]. Video super-resolution can be considered as an extension of image SR by the incorporation of recursive estimation approach. Video SR aims at providing high quality resolution enhancement for videos. A fast and adaptive video super-resolution algorithm is always desirable and is also the trend for future research.

According to [2], there are two groups of video SR algorithms: learning- and reconstruction-based. Learning-based algorithms build a database of image pairs; each pair consists of low resolution and high resolution patches. The low resolution image is enhanced by applying the high resolution patches onto the low resolution image. Reconstruction-based algorithms make use of several input frames to reconstruct the high resolution image by registration, fusion and restoration processes. More details can refer to [3], [4].

According to [5], video SR can also be classified as static and dynamic. Static SR does not make use of information from previously super-resolved frame, and thus cannot provide the temporal consistency of enlarged video. Dynamic SR like [6] and [7] are dependent of previously reconstructed frame.

In this paper, we focus on fast dynamic reconstruction-based video SR. For such type of fast algorithms, it is assumed that the previous frame is well reconstructed. The previously reconstructed frame acts only as a reference frame for motion estimation per block, such that the registration process becomes a motion compensation process. Due to the inaccuracy of motion compensation, the registration error could be large, except for purely translational motion.

If the motion compensation accuracy is significantly increased, the registration error is decreased accordingly. In this paper, we propose a novel idea that the compensation accuracy can be increased by considering the high frequency components in previous reconstructed frame as the only compensated parts to be applied to the current up-sampled observed frame. In [8], Freeman, Jones and Pasztor confirmed the feasibility of frequency separation of image patches. [6] Bishop, Blake and Marthi further extended the idea into a dynamic video SR algorithm. We integrate the idea of frequency classification and motion compensation to form a new motion compensation model for use in typical dynamic SR framework. The most important part of our idea is that it avoids the significantly varying low frequency components of pixels between frames, such that registration error is only influenced by the high frequency components.

In section 2, we will present our model and the SR framework. Experimental results in Section 3 show that the new translational motion compensation model obviously increases the objective and subjective quality of the motion compensated frame. With the new motion compensated frame, we are able to extend from two frames fusion to the three frames fusion. Experimental results show that three-frame fusion can increase objective quality, compared with the two-frame fusion. Section 4 concludes the paper.

2. ALGORITHM

Typical fast dynamic Video SR Error minimization model can be represented as a modification of the typical image SR model by the incorporation of recursive estimation approach, as shown below:

$$z_R = \arg\min_z \sum_i \left\| A_{k-i} z_{k-i} - w_k \right\| \quad (1)$$

where $A_{k-1}z_{k-1}$ is the ith simulated current Low Resolution (LR) frame from the ith previously reconstructed High Resolution (HR) frame z_{k-1} and $A_{k-i}z_{k-i} = DH_{cam}F_{k-i}H_{atm}z_{k-i} + n$. H_{atm} is the turbulence effect that is ignored due to its insignificance in value. F_{k-i} is the warping operator. H_{cam} is the camera lens blur, for which its convolution operator is equal to $\frac{1}{4}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$. D is the decimation operator. Let us consider a noiseless environment, therefore n as the noise term can be ignored. w_k is the observed current LR frame.

In this model, we aim at minimizing the registration error given a warping operator F_{k-i}. Let us just consider block-based approach with translational motion, due to the sake of simplicity and computational cost. In case of one previous reconstructed frame is used, i.e. i=1, the model is simplified to:

$$z_R = \arg\min_z \left\| A_{k-1} z_{k-1} - w_k \right\| \quad (2)$$

where $A_{k-1}z_{k-1} = DH_{cam}F_{k-1}H_{atm}z_{k-1} + n$. It is assumed that $w_k = DH_{cam}z_{gt}$, where z_{gt} is ground true HR current frame. The corrected error model by considering the ignored H_{atm} becomes: $z_R = \arg\min_z \left\| DH_{cam}F_{k-1}z_{k-1} - DH_{cam}z_{gt} \right\|$ or

$$z_R = \arg\min_z \left\| DH_{cam}F_{k-1}z_{k-1} - w_k \right\| \quad (3)$$

We propose that traditional motion compensated frame, $F_{k-1}z_{k-1}$, be replaced by $F_{k-1}z_{k-1}(highfreq)+Uw_k$ in this paper, such that the registration error is reduced. Hence the new error model becomes:

$$z_R = \arg\min_z \left\| DH_{cam}(F_{k-1}z_{k-1}(high.freq) + Uw_k) - w_k \right\| \quad (4)$$

In order to further minimize the error function in (4), a fusion process is carried out, as shown in Section 2.3.

2.1. Frequency extraction and separation

According to [6] and [8], a block of pixels containing full spectrum of frequency can be separated into high and low frequency components by low-passing filtering, down-sampling and subtraction. Our approach making use of frequency extraction and separation bears some similarities with the approaches in [6] and [8]. However, they are different. First, our approach does not normalize the contrast of the low frequency and high frequency components due to our desire of using approximately consistent contrast between frames. Second, we use an optimal up-sample operator (Table 1), rather than the Cubic Spline as in [6] and [8]. Third, we use a camera lens blur as the low pass operator, which is different from the operator in [6]. Mathematically, our method is described below. The way to obtain the low frequency component $z_{k-1}(low.freq)$ from the previously reconstructed frame z_{k-1} as shown below:

$$z_{k-1}(low.freq) = UDH_{cam}z_{k-1} \quad (5)$$

where D is the decimation operator and U is the up-sampling operator. Let us assume perfect separation of frequency components. The high frequency component $z_{k-1}(high.freq)$ can be obtained by:

$$z_{k-1}(high.freq) = z_{k-1} - z_{k-1}(low.freq) \quad (6)$$

We use the same decimation operator and camera lens blur as in Section 2, and lanczos4 as the U operator. Lanczos4 is the windowed Sinc function that will approximate closely the optimal re-sampling filter than any other linear methods. Again, lanczos4 is used as the up-sample operator $\boldsymbol{U}$ to obtain the up-sampled frame Uw_k from observed input fr. w_k.

2.2. New motion compensation

Recall the traditional motion compensation of $F_{k-1}z_{k-1}$, which is to be replaced by the new model, $F_{k-1}z_{k-1}(highfreq)+Uw_k$. We will show that the new model is able to give less registration error compared with the traditional model.

In this work, the block-based motion estimation without sub-pixel precision is used. This assumes translational motion activities only. A set of motion vectors, d_{block}, is obtained, one for each block, between the reference frame z_{k-1} and the up-sampled current frame Uw_k making use of the SAD (Sum of absolute difference):

$$d_{block} = \mathrm{SAD}\,(z_{k-1}, Uw_k) \quad (7)$$

We consider a block size of 16x16 and a search window of 32x32. The SAD framework can be optimized using pixel decimation[9], with only a quarter of the pixels extracted for calculating sum of absolute difference. The motion vectors, d_{block}, are then used to define the addresses for motion compensation of high frequency blocks from the frame of high frequency components $z_{k-1}(highfreq)$ that we obtained earlier in(6):

$$F_{k-1}z_{k-1}(high.freq) + Uw_k = z_{k-1}(high.freq)(d_{block}) + Uw_k \quad (8)$$

where $F_{k-1}z_{k-1}(highfreq)+Uw_k$ is the motion compensated frame by combining the high frequency part of previous reconstructed frame and the up-sampled current frame. The registration process is then changed and registration error is only influenced by the high frequency components. For traditional full spectrum range of motion compensation:

$$F_{k-1}z_{k-1} = z_{k-1}(d_{block}) \quad (9)$$

The difference between the computational costs of obtaining $F_{k-1}z_{k-1}(highfreq)+Uw_k$ and $F_{k-1}z_{k-1}$ is negligible, since $F_{k-1}z_{k-1}(highfreq)+Uw_k$ involves only one additional step (6). Moreover, we have investigated the optimal up-sample operators U for our model. As shown in Table 1, using lanczos4 as U operator gives the highest PSNR of $F_{k-1}z_{k-1}(highfreq)+Uw_k$, and thus lanczos4 is chosen.

Table 1 PSNR of new motion compensated frame using different up-sample operators U

	Lanczos4	Edge-directed [7]	Bilinear
PSNR(dB)	25.76	25.29	24.78

2.3. Fusion process

Fusion is a process to fuse several frames together, such that the overall error in (4) can be further minimized. Different from the fusion process in [7], the calculation of fusion coefficients in our approach is not done by training-set and is straightforward, rather than using AdaBoost classifier [10].

Note that a three-frame fusion is better than a two-frame fusion because of additional information that the fusion model can make use of. Experimental results of the fusion are shown in Table 2. As shown in (10), there are three available frames we want to fuse together, they are the up-sampled current frame, Uw_k, the traditional motion compensated frame, $F_{k-1}z_{k-1}$, and the new motion compensated frame, $F_{k-1}z_{k-1}(highfreq)+Uw_k$. The fusion equation for fusing three frames is as shown below:

$$z_R = \arg\min_z \left\| \begin{array}{l} Coe_1 DH_{cam}(F_{k-1}z_{k-1}(high.freq) + Uw_k) + \\ Coe_2 DH_{cam} F_{k-1}z_{k-1} + Coe_3 DH_{cam} Uw_k - w_k \end{array} \right\|$$
$$Coe_1 + Coe_2 + Coe_3 = 1 \qquad (10)$$

and Coe_i is the fusion coefficient of ith fusing frame for $(x,y) \in w_k$. Coe_i are calculated by the normalization equation, as shown below:

$$Coe_i = \frac{\frac{\sum_{j=1,2,3}(e_j)}{e_i}}{\sum_{k=1,2,3}\left\{\frac{\sum_{j=1,2,3}(e_j)}{e_k}\right\}} \quad \text{for } i=1,2,3 \qquad (11)$$

where e_i is squared error of ith fusing frame derived from:

$$e_1 = (DH_{cam}(F_{k-1}z_{k-1}(high.freq) + Uw_k) - w_k)^2;$$
$$e_2 = (DH_{cam}F_{k-1}z_{k-1} - w_k)^2; e_3 = (DH_{cam}Uw_k - w_k)^2 \quad (12)$$

for $(x,y) \in w_k$. Each e_i is added by 1 to avoid division by zero. Eventually, the calculated fusion coefficients Coe_i represent the weighting factors of four pixels in the HR grid for the ith fusing frame, such that:

$$z_k(2x+l, 2y+k) = \sum_{i=1,2,3} Coe_i H_i(2x+l, 2y+k) \qquad (13)$$

for $l,k=0,1$ and $(x,y) \in w_k$. H_i represents the three fusing frames and z_k is the final reconstructed current frame. To fuse two frames, the method is straightforward by considering one less component in (10), (11), (12) and (13), and i=1,2. Note that our approach can support infinite number of fusing frames theoretically.

3. EXPERIMENTAL RESULTS

The tests were run on an Intel 3G dual core system. Six video sequences of resolution 1280x720 were down-sampled and then up-sampled. The sequences include global and local motions, and are not constrained to translational motion. An initial estimation z_0 is obtained by interpolation using lanczos4. The sequences are in YUV format. All the experiments were conducted using Y component only, except for Fig. 4. In this latter case, we processed the three channels of YUV format independently for displaying the color comparison. The computational times per frame of the Bilinear, Lanczos4, motion compensation and the proposed approach (including fusion, etc.) are, on average, 0.02s, 0.05s, 2.8s and 3.1s respectively by using non-optimized C++ codes.

Let us verify that the new motion compensated frame, $F_{k-1}z_{k-1}(highfreq)+Uw_k$, is better than the traditional motion compensated frame, $F_{k-1}z_{k-1}$. Brandi *et al* provided a comparable approach to motion compensation through frequency classification [12]. As shown in Table 2, the PSNR result of the new motion compensation model is on average 2dB higher than the traditional model and 1dB higher than the motion compensation method in [12]. As shown in the Ducks Take Off sequence (Fig.1), when the motion compensation accuracy decays, the PSNR of the traditional motion compensation decreases rapidly. In contrast, the PSNR of the new model is close to or even better than that of lanczos4. Fig 2 shows the analysis of the PSNR of sequence 1. Moreover, the absolute residual error of the new motion compensation model is obviously less than that of the old compensation model in Fig 3. These results agree to our new model which reduces the registration error under the assumption in Section 2. In conclusion, the subjective and objective qualities of the new motion compensation model are better than those of the traditional model.

Let us show that the fusion process in Section 2.3 can reduce the overall error in (4), and thus increase the SR quality by making use of the new motion compensated frame. As shown in Table 2, the PSNR resulting from three frame fusion approach is on the average 0.25dB higher than those of the two frame fusion of lanczos4 and old motion compensated approach. This confirms the contribution of the new motion compensated frame to the SR quality after the fusion process. Note also that the PSNR resulting from three frame fusion is on the average 1.5dB and 3dB higher than that of Lanczos4 and Bilinear methods respectively, as shown in Table 2. The SR quality of the three frame fusion provides a sharper and more detailed picture than the Lanczos4 and Bilinear as shown in Fig 4. This agrees that the SR quality is better than the linear methods.

4. CONCLUSION

In this paper, we have presented a new motion compensation model and a SR framework that effectively reduce the error in traditional model for super-resolution enhanced videos. The computational cost of the proposed method is moderate and is suitable for real-time and near real-time applications. 80% of computational time goes to motion compensation. By further optimizing the motion compensation model, the computational time would possibly be reduced, which is a direction of our future work. Experimental results show that the subjective and objective qualities of the new model that is suitable for SR framework are better than that of the traditional model. The SR quality would further be improved if a more sophisticated fusion process is applied. Hence, a

future research direction is to look for a new fusion process that can further address the error minimization problem.

Acknowledgement: This work is supported by the Center for Signal Processing, the Hong Kong Polytechnic University and the Research Grant Council of the Hong Kong SAR Government (PolyU 5278/08E).

Table 2 Y-PSNR (dB) for various methods

Method	Video Sequences					
	1	2	3	4	5	6
(1)Bilinear	**23.75**	**23.43**	**26.07**	**26.67**	**28.88**	**28.13**
(2)Lanczos4	**25.57**	**24.79**	**28.37**	**28.23**	**29.78**	**29.65**
(3)Old Motion compen.	**23.51**	**22.81**	**25.77**	**27.5**	**31.07**	**29.35**
(4)New motion compen.	**26.09**	**25.27**	**28.47**	**29.38**	**31.79**	**30.38**
Motion compen. [12]	**25.14**	**24.54**	**27.78**	**27.77**	**29.46**	**29.23**
Fusion of (2) and (3)	**26.34**	**25.45**	**29.15**	**29.83**	**32.41**	**31.06**
Fusion of (2) and (4)	**26.35**	**25.5**	**29.06**	**29.72**	**31.99**	**30.85**
Fusion of (2), (3) and (4)	**26.56**	**25.65**	**29.4**	**30.05**	**32.65**	**31.31**
Max-Likelihood[11] on (3)	**26.53**	**25.73**	**28.98**	**30.13**	**33.02**	**29.36**
Max-Likelihood[11] on (4)	**26.58**	**25.69**	**29.17**	**29.91**	**32.4**	**31.06**

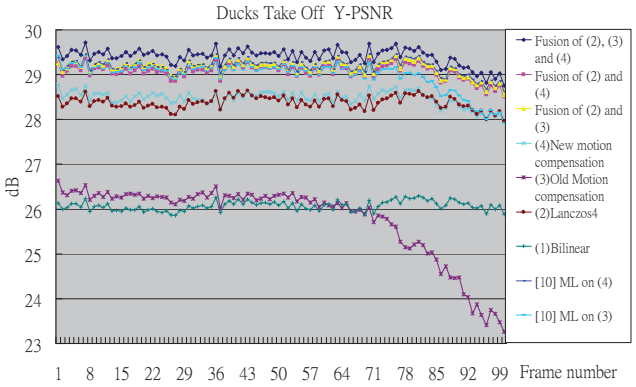

FIG.1: Y-PSNR(DB) OF SEQUENCE 3 (IN TABLE2)

5. REFRENCES

[1] S.C. Park, M.K. Park, and M.G. Kang, "Super-Resolution Image Reconstruction: A Technical Overview", IEEE Signal Processing Magazine, Vol. 20, pp. 21-36, May 2003.

[2] A. Krylov and A. Nasonov, "Fast super-resolution from video data using optical flow estimation", ICSP'2008, 9th Int. Conf. Signal Processing, vol., no., pp.853-856, 26-29 Oct. 2008.

[3] S. Baker and T.Kanade, "Limtis on Super-Resolution and How to Break Them," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.24, Issue 9, pp. 1167-1183, Sept. 2002.

[4] B.K. Gunturk, A.U. Batur, Y. Altunbasak, M.H. Hayes III; R.M. Mersereau, "Eigenface-domain super-resolution for face recognition," Image Processing, IEEE Transactions on , vol.12, no.5, pp. 597-606, May 2003.

[5] S. Farsiu, M. Elad, and P. Milanfar, "Video-to-Video Dynamic Super-Resolution for Grayscale and Color Sequences", EURASIP Journal on Applied Signal Processing, No. Article ID 61859, pp.1-15, 2006.

[6] C. Bishop, A. Blake, and B. Marthi, "Super-resolution enhancement of video", in C. M. Bishop and B. Frey (Eds.), Proc. of the Ninth International Workshop on Artificial Intelligence and Statistics, January 2003.

[7] Simonyan, K.; Grishin, S.; Vatolin, D.; Popov, D., "Fast video super-resolution via classification," ICIP'2008, 15th IEEE International Conference on Image Processing, pp.349-352, 12-15 Oct. 2008.

[8] W. T. Freeman, T. R. Jones, and E. C. Pasztor. "Example-based superresolution", IEEE Computer Graphics and Applications, vol.22(2), pp. 56–65, March/April 2002.

[9]Yiu-Lam Chan and Wan-Chi Siu, 'New Adaptive Pixel Decimation for Block Motion Vector Estimation', IEEE Transactions on Circuits & Systems for Video Technology, pp.113-118, Vol.6, No.1, Feb., 1996.

[10] J. H. Friedman, T., Hastie, and R. Tibshirani, "Additive logistic regression: a stat. view of boosting", Dept. of Statistics, Standford U, Tech Rpt., 1998.

[11]M. Irani and S. Peleg, "Improving resolution by image registration," CVGIP: Graph. Models and Image Proc., vol. 53, pp.231-9, May 1991.

[12] F. Brandi, R. de Queiroz, and D. Mukherjee, "Super-resolution of video using key frames and motion estimation", pp.321-324, Proc., 15th IEEE International Conference on Image Processing, 12-15 Oct. 2008.

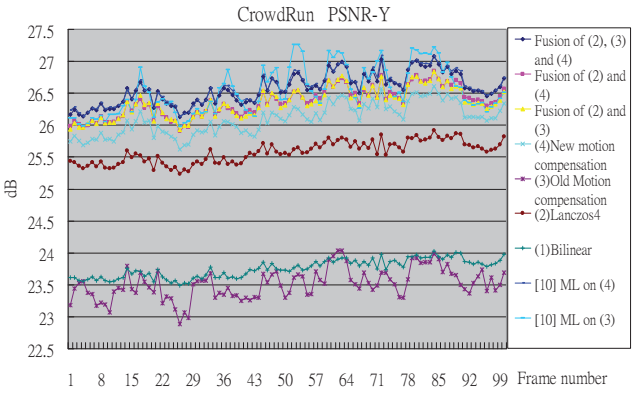

Fig.2: Y-PSNR(dB) of sequence 1 (in table2)

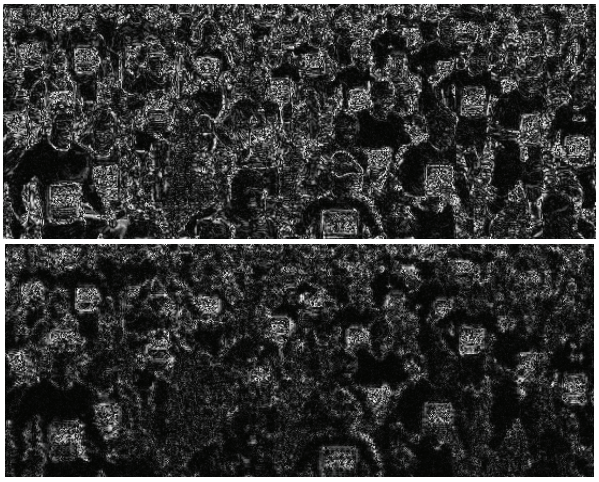

Fig.3: Portions of absolute residual error. Above and below figures are the traditional and new motion compensated frames.

Fig.4: Portions of sequence 5. Left figure is Bilinear. Middle figure is Lanczos4. Right figure is SR-three-frame fusion.