A Novel Hybrid Technique in Spatio-Frequecncy Domain for Better Quality Image and Video Up-Sampling

Sobhan Kanti Dhara
Dept. of Electronics and Communication Engg.
National Institute of Technology Rourkela
Rourkela-769008, India
dhara.sk@gmail.com

Sukadev Meher
Dept. of Electronics and Communication Engg.
National Institute of Technology Rourkela
Rourkela-769008, India
sukadevmeher@gmail.com

Abstract— **Generally, at the transmitter video intra frame is sub-sampled by resizing the frame. It is very important to generate good quality high resolution frame from the receiver low resolution frame. The existing most of all up-sampling technique produce undesirable blurring and ringing artefacts. To overcome this problem, a novel technique is proposed here which use both space and frequency domain information. The proposed method use low frequency DCT (Discrete cosine transform) component to sub-sample the frame or image at the transmitter side. In transmitter side a preprocessing method is proposed where the received sub-sampled frame or image is passed though a Wiener filter which uses its local statistics in 3×3 neighbourhood to modify pixel values. The output of Wiener filter is added with optimized multiple of high frequency component. The output is then passed through a DCT block to up-sample. Result shows that the proposed method outperforms most of the widely used existing interpolation techniques in terms of objective and subjective measures.**

Keywords- Discrete cosine transform; interpolation; local variance ; ringing artifact; up-sampling

I. Introduction

Image up-sampling is an important technique to produce high resolution image or frame at receiver side from a low resolution received image or frame from the transmitter side at the receiver end. Generally, at the transmitting end, a video intra-frame or image is down-sampled to lessen the bandwidth required for transmission and to avoid channel congestion. At the receiver, low resolution of the down-sampled intra-frame or image is up-sampled to its original resolution by different interpolation techniques. These interpolation techniques are useful for displaying high definition standard display. In addition, interpolation plays a significant role in applications like medical diagnosis, satellite image monitoring, video surveillance and many more. In such applications, it is very often required to improve the native resolution of the original image for proper inspection and recognition [1].

There are many interpolation techniques used to up-sample the low resolution frame or image. The simple interpolation techniques among them are bilinear interpolation where the output pixel value is a weighted average of pixels in the nearest 2-by-2 neighborhood [2].

Though simple, Bilinear interpolation has undesirable blurring artefacts. There are other common used interpolation techniques [3, 5] such as Bicubic where the output pixel value is interpolated by weighted average of pixels in the nearest 4-by-4 neighborhood. Bicubic interpolation technique has a less blurring in compare to linear interpolation. Lanczos is another spatial domain interpolation technique which is implemented by multiplying a sinc function with a sinc window which is scaled to be wider and truncated to zero outside of a defined range [7]. Even if Lanczos interpolation gives good results, it is slower than other approaches and provides a blurring artefact in the reconstructed image. Up-sampling in DCT domain is implemented by padding zero coefficients to the high frequency side. Image resizing in DCT domain shows very good result in terms of scalability and image quality. However, this technique also suffers through undesirable blurring and ringing artefacts [1].

Thus, it shows that different interpolation techniques produce unnecessarily blurring and ringing artefacts while up-sampling. So, the requirement is to have good subjective and objective quality up-sampled frame or image having very less amount of blurring and ringing artefacts. The proposed hybrid method process the down-sampled image both in spatial and frequency domain to meet the requirements.

The organization of the paper is structured as follows. The proposed method is described in the subsequent section. Section-III provides the simulation results of different interpolation algorithms subjected to various constraints. Finally, the work is concluded in section-IV.

II. Proposed Method

In transmitter side instead of alternate deletion of rows and columns, a sub-sampled video intra-frame or image is produced by taking 25% of DCT components residing upper left corner of DCT output of image of video intra-frame. The output of IDCT of the mentioned DCT coefficients gives a low resolution image size half of the actual. At the receiver, this low resolution sub-sampled video or image is to be up-sampled with a good objective and subjective quality. Up-

978-1-4799-5364-6/14/$31.00 ©2014 IEEE

sampling using interpolation is similar to low pass filter operation and like any low pass filtering operation it also suffers from blurring effects at the filtered output. To overcome this, spatial domain statistical approach is taken by passing the input to Wiener filter. It uses its 3×3 neighborhood to update the pixel value [11]. It estimates the local mean and variance around each pixel.

A novel spatial domain preprocessing technique is proposed using DCT interpolation technique and Lanczos 3 interpolation to get less blurring and retain some fine details and edge information.

A. Down-sampling in the DCT domain

To implement down-sampling in DCT domain, we take DCT of 2N×2N image. Then we take IDCT of upper left N×N DCT coefficients to make it N×N image or intra frame.

B. Wiener Filter based pre processing

Here, in the receiver spatial domain statistical approach is taken by passing the input to Wiener filter. It uses its 3x3 neighborhood to update the pixel value [11].It estimates the local mean and variance around each pixel as μ and σ^2 respectively.

Wiener filter then creates a pixel wise filter using following estimates,

$$b(n1,n2) = \mu + \frac{\sigma^2 - v^2}{\sigma^2}(a(n1,n2) - \mu) \tag{1}$$

where, v^2 is the average of all the local estimated variance.

A spatial domain preprocessing technique is approached to get less blurring and retain some fine details and edge information. In general some high frequency component is added as to compensate the loss of edge information for typical sharpening or high boosting operation. Instead of this typical sharpening here we first extract the high frequency components from the wiener filter output. Then we propose here to add some weighted factor K times extracted high frequency component with the Wiener output. The weight factor, K is determined adaptively for the first sub-sampled frame or by each frame referring the corresponding original frame or image [11].

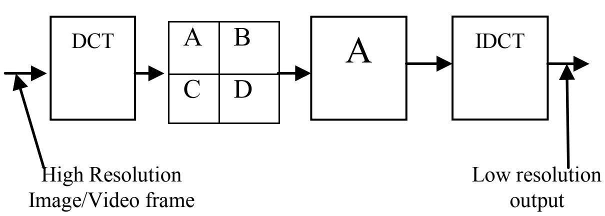


Figure 1. DCT based down-sampling in Transmitter side

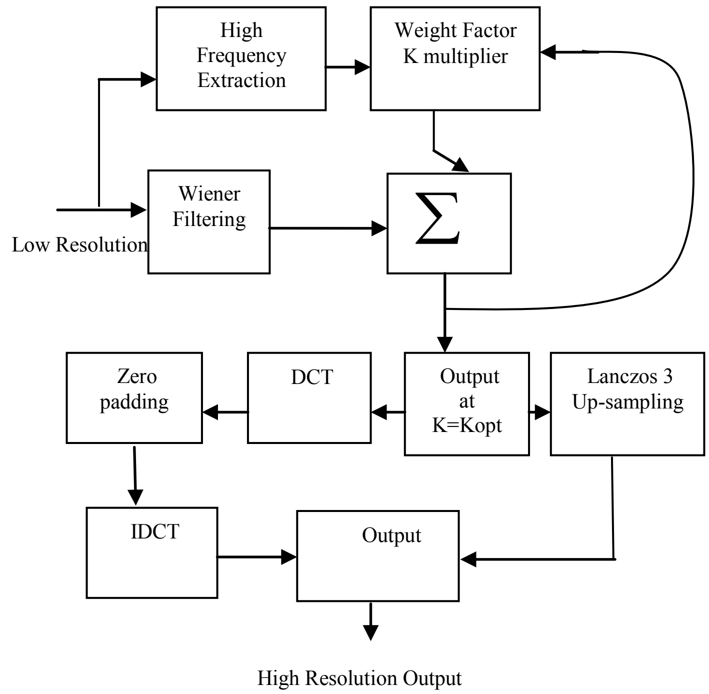


Figure 2. Reiciver side Frame/Image high resolution output

C. Up-sampling in the DCT domain and Lanczos 3 domain

To implement up-sampling in DCT domain, we need to add N zeros in the high frequency regions, where N is the signal length. After that, type-II IDCT of the extended 2N samples is performed to obtain the two fold up-sampled data [7]. In case of video frames or images, the up-sampling in a matrix form is given by

$$b^U_{2N\times 2N} = W^T_{2N\times 2N} \times \begin{pmatrix} 2W_{N\times N} b_{N\times N} W^T_{N\times N} & 0 \\ 0 & 0 \end{pmatrix} \times W_{2N\times 2N} \tag{2}$$

where W denotes the 1-D type-II DCT kernel. b and b^U are the down-sized and the up-sampled frame block. 0 is N×N zero matrix [8].

Lanczos is a spatial domain interpolation technique which is implemented by multiplying a sinc function with a sinc window which is scaled to be wider and truncated to zero outside of the main lobe. In case of Lanczos-3 interpolation, the main lobe of the sinc function along with the two subsequent side lobes on either side is used as a sinc window [1, 7].

The pixel on the interpolated values is defined by the filter's Lanczos kernel *L(x)*. The Lanczos window is the normalized sinc function sinc(x), multiplied restricted to the main period $-a \le x \le a$ *to* form a convolution kernel for re-sampling the input field [7].

$$L(x)=\begin{cases}\sin c(x)\sin c(x/a), -a\le x\le a\\ 0, Otherwise\end{cases} \quad (3)$$

III. Parameter to be Measured

The following parameters are taken into consideration to measure the amount of denoising. We consider the restored image as f(x, y) and original image as g(x, y).

Peak Signal to Noise Ratio (PSNR)

It is peak signal to noise ratio where R mentioned as the maximum fluctuation of input image and MSE is the mean square error. Mathematically it is expressed as

$$PSNR = R^2 / MSE \quad (4)$$

All the above mentioned parameters are quantitative measure of denoising and do not relate to perceived image quality.

Structural Similarity Index Metric (SSIM)

The Structural Similarity (SSIM) which is one of the quantitative metric measure the quality of an image. It compares between a reference image and a distorted image. It consist luminance comparison, contrast comparison and structural comparison. First, the luminance of each signal is compared. Assuming discrete signals, this is estimated as the mean intensity:

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (5)$$

Second, we remove the mean intensity from the signal.

$$\sum_{i=1}^{N} x_i = 0 \quad (6)$$

We use the standard deviation (the square root of variance) as an estimate of the signal contrast.

$$\sigma_x = (\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2)^{1/2} \quad (7)$$

The contrast comparison c(x; y) is then the comparison of standard deviation of x and y.
For luminance comparison, we define

$$l(x,y)=\frac{2\mu_x\mu_y + c_1}{{\mu_x}^2+{\mu_x}^2+c_2} \quad (8)$$

Where L is the dynamic range of the pixel values (255 for 8-bit grayscale images).The contrast comparison function takes a similar form that end the value

$$c(x,y)=\frac{2\sigma_x\sigma_y + c_2}{{\sigma_x}^2+{\sigma_y}^2+c_2} \quad (9)$$

$$c_2=(k_2L)^2 \quad (10)$$

The structure comparison functions as follows:

$$s(x,y)=\frac{2\sigma_x\sigma_y + c_3}{\sigma_x\sigma_y+c_3} \quad (11)$$

We define SSIM as

$$SSIM(x,y)=[l(x,y)]^{\alpha}.[c(x,y)]^{\beta}.[s(x,y)]^{\gamma} \quad (12)$$

IV. Experimental Results and Discussion

To compare the performance of the proposed post-processing scheme, we have taken some 11 test videos sequence and 5 images for input test signals. The inputs are down-sampled in the spatial domain by resizing 4:1 compression ratio and another case sampled by DCT method mentioned above. We up-sampled the frames back to their original resolution to compare with the original video frame or images. Table I and Table II illustrate the average PSNR and SSIM comparison of DCT, Lanczos-3, Fuzzy Unsharp [12] Proposed I and Proposed II techniques for test images. Table III and Table IV illustrate the average PSNR and SSIM comparison of DCT, Lanczos-3, Fuzzy Unsharp [12] Proposed I and Proposed II techniques for test video sequences. In each case, the original image or frame down-sampled at 4:1 ratio. So the encoded bits of the frame or image at the output for the all mentioned algorithms are almost same. Thus, with respect to compression ratio or output encoded bit streams (Kb/Mb), all the algorithm has almost similar efficiency. Lancozos-3 and DCT based algorithms have less computational complexity respect to other two. Fuzzy based technique has an extra overhead of Fuzzy implementation and K optimization. But proposed algorithms have only the K optimization overhead in terms of computational complexity. Taking into consideration this, we analyzed our proposed algorithm and it shows that it performs better with respect to other algorithm mentioned for comparison. Experimental results reveal that for both images and video sequences Proposed II is giving best result among the four techniques with respect to PSNR and SSIM. Proposed I method is also giving better result for the most of the cases with respect to the other three methods. We have shown here PSNR comparison for two video sequences Akiyo and Foreman in the graph for some defined number of frames in figure 3. It clearly shows our proposed algorithm is giving the best result. The subjective performance of the proposed technique is illustrated in Figure 4 and Figure 5 for test image Lena and the 25th frame of akiyo sequence at 4:1 compression ratio correspondingly.

TABLE I. Average PSNR Comparison of Different Image at 4:1 Compression Ratio

Video sequences	**PSNR (dB)**				
	Lanczos 3	***DCT***	***Fuzzy-Unsharp [11],[12]***	***Proposed I***	***Proposed II***
Lena	34.7931	35.0010	35.1741	35.3526	**36.0014**
Pepper 512	32.0867	32.5439	32.5655	32.3857	**33.4933**
Boat	30.3747	30.4661	30.6089	30.7053	**31.2170**
Baboon	23.8583	23.9234	24.0316	24.0764	**24.5123**
City	25.4280	25.1834	25.1545	25.4268	**25.6923**

TABLE II. AVERAGE SSIM COMPARISON OF DIFFERENT IMAGE AT 4:1 COMPRESSION RATIO

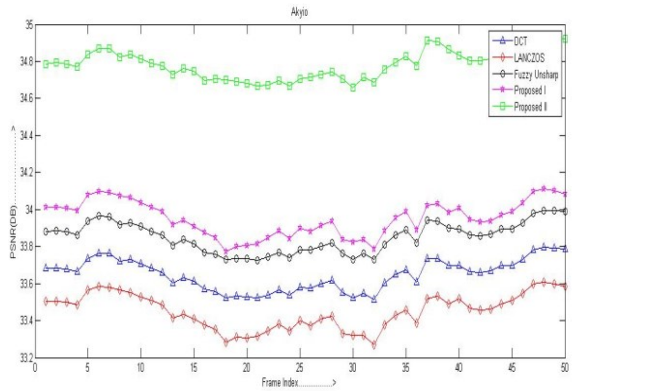
Video sequences	Average SSIM				
	Lanczos 3	*DCT*	*Fuzzy-Unsharp [11],[12]*	*Proposed I*	*Proposed II*
Lena	0.9953	0.9955	0.9957	0.9959	**0.9965**
Pepper 512	0.9831	0.9838	0.9839	0.9935	**0.9950**
Boat	0.9862	0.9865	0.9871	0.9873	**0.9887**
Baboon	0.9185	0.9201	0.9237	0.9237	**0.9334**
City	0.9679	0.9661	0.9662	0.9681	**0.9701**

TABLE III. AVERAGE PSNR COMPARISON OF DIFFERENT SEQUENCES AT 4:1 COMPRESSION RATIO

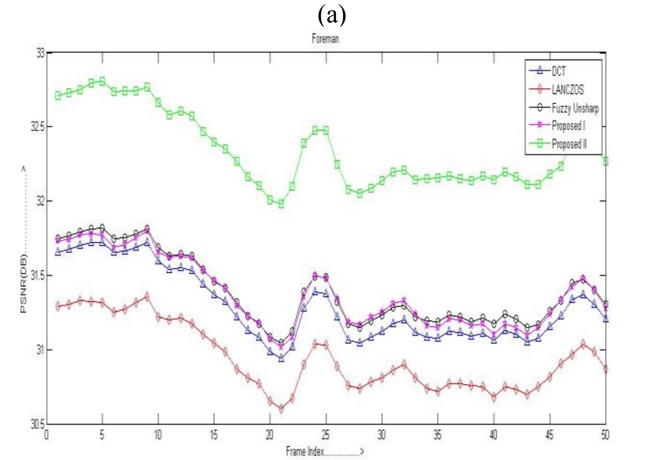
Video sequences	Average PSNR (dB)				
	Lanczos 3	*DCT*	*Fuzzy-Unsharp [11],[12]*	*Proposed I*	*Proposed II*
Akiyo	33.4501	33.6473	33.8500	33.9526	**34.7824**
Bus	24.4694	24.5392	24.7582	24.8548	**25.6156**
City	27.8791	27.8520	27.9551	28.1085	**28.6690**
Coastguard	26.9391	27.0806	27.2918	27.3208	**28.2468**
Container	26.0086	26.2563	26.4553	26.4066	**27.432**
Flower	22.0333	22.0957	22.1774	22.2725	**22.7767**
Football	29.3665	29.687	30.0840	30.1332	**31.2577**
Foreman	30.9453	31.2917	31.3927	31.3762	**32.3415**
Hall_monitor	26.8520	27.2858	27.4766	27.1846	**28.6440**
Ice	33.1415	33.2192	33.0547	33.5799	**34.2834**
Mobile	21.5967	21.7580	21.9851	22.0403	**22.7815**

TABLE IV. AVERAGE SSIM COMPARISON OF DIFFERENT SEQUENCES AT 4:1 COMPRESSION RATIO

Video sequences	Average SSIM				
	Lanczos 3	*DCT*	*Fuzzy-Unsharp [11],[12]*	*Proposed I*	*Proposed II*
Akiyo	0.99531	0.99553	0.99575	0.99584	**0.996576**
Bus	0.96059	0.96188	0.96495	0.96475	**0.970512**
City	0.93927	0.93871	0.94169	0.94342	**0.950972**
Coastguard	0.97453	0.97541	0.97694	0.97689	**0.981835**
Container	0.97139	0.97318	0.97489	0.97441	**0.980046**
Flower	0.94339	0.94439	0.94709	0.94754	**0.953626**
Football	0.98022	0.98182	0.98387	0.98351	**0.987752**
Foreman	0.99314	0.99365	0.99385	0.99379	**0.9950**
Hall_monitor	0.97738	0.97979	0.98097	0.97944	**0.98559**
Ice	0.99416	0.99432	0.99412	0.99474	**0.99563**
Mobile	0.95657	0.95849	0.96123	0.96156	**0.96826**



(a)



(b)

Figure 3. PSNR comparison at 4:1 compression ratio: (a) akiyo; (b) Foreman .

(a)

(b)

(c)

(d)

Figure 4. Subjective performance of Test Image Lena at 4:1 compression ratio: (a) Original; (b) Lanczos-3; (c) DCT; (d) Fuzzy Unsharp (e) Proposed I ;(f) Proposed II

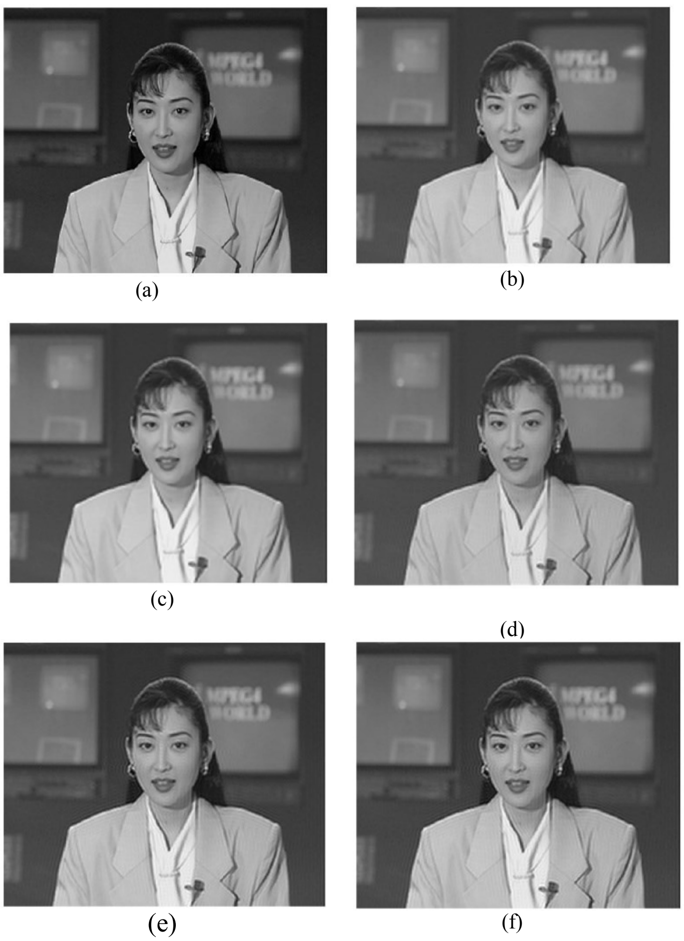


Figure 5. Subjective performance of 25th frame of akiyo sequence at 4:1 compression ratio: (a) Original; (b) Lanczos-3; (c) DCT; (d) Fuzzy Unsharp ;(e)Proposed I ;(f) Proposed II

V. Conclusion

The proposed schemes are an approach to restore the lost information during the down sampling and up-sampling operation. The computational overhead can be minimized through fast optimization algorithms. It restores the fine details and edge information of video and image. It alleviates the problem of blurring artefacts. The improvement is gained by using the local statistics based Wiener filtering with statistical local variance of a neighbourhood on direct mapping basis. In addition, the proposed scheme performs quite adaptively under various constraints such as change in compression ratio (i.e. 16:1). This method adaptively removes noise content added to the image or video frame. Moreover, the good subjective quality can be visualized in terms of sharpened edge; less degree of blurring and fine details preservation. Thus this algorithm is giving good qualitative and subjective quality output.

References

[1] Acharya, Aditya, and Sukadev Meher. "Local Adaptive Laplacian for Better 2-D Up-sampling." 2013 2nd International Symposium on Computer, Communication, Control and Automation (3CA 2013), December 1-2, 2013, Singapore. 2013.

[2] Lu Jing, Xiong Si, Wu Shihong, “An improved bilinear interpolation algorithm of converting standard defination images to high defination images,” WASE Int. Conf. on Info. Engg. pp.441-444, 2009.

[3] R. G. Keys, “Cubic convolution interpolation for digital image processing,” IEEE Trans. Acoust., speech, signal Process., vol. ASSP-29, no.6, pp.1153-1160, Dec.1981.

[4] S. E. Reichenbach and F.Geng, “Two-dimensional cubic convolution,” IEEE Trans. Image Process., vol.12, no.8, pp.857-865, Aug. 2003.

[5] Zhou Dengwen, “An edge directed bicubic interpolation algorithm,” CISP, pp.1186-1189, 2010.

[6] Hou, Hsieh S., and H. Andrews. "Cubic splines for image interpolation and digital filtering." Acoustics, Speech and Signal Processing, IEEE Transactions on 26.6 (1978): 508-517.

[7] Wenxing Ye, Alireza Entezari, A geometric construction of multivariate sinc functions, IEEE Transaction on Image processing 2011; 19(12).

[8] R. Dugad and N. Ahuja, “A fast scheme for image size change in the compressed domain,” IEEE Trans. Circuit, Syst., Video Technology., vol. 11, pp. 461-474, Apr, 2001.

[9] Zhenyu Wu, Hongyang Yu, and Chang Wen Chen, “A new hybrid DCT-Wiener based interpolation scheme for video intraframe up-sampling,” *IEEE* signal processing letters, vol. 17. No. 10, pp. 827-830, oct. 2010.

[10] Lim, Jae S., Two-Dimensional Signal and Image Processing, Englewood Cliffs, NJ, Prentice Hall, 1990, p. 548, equations 9.44 -- 9.46.

[11] Acharya, Aditya, and Sukadev Meher. "An efficient, adaptive unsharp masking based interpolation for video intra frame up-sampling." Microelectronics and Electronics (PrimeAsia), 2012 Asia Pacific Conference on Postgraduate Research in. IEEE, Dec.2012.

[12] Acharya, Aditya, and Sukadev Meher. "No reference, fuzzy weighted unsharp masking based DCT interpolation for better 2-D up-sampling." Fuzzy Systems (FUZZ), 2013 IEEE International Conference on. IEEE, 2013.