An Efficient, Adaptive Unsharp Masking Based Interpolation for Video Intra Frame Up-sampling

Aditya Acharya
Dept. of Electronics and Communication Engg.
National Institute of Technology
Rourkela, 769008, India
aditya.acharya2011@gmail.com

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

Abstract—**Most of the existing interpolation techniques available in the literature provide an undesirable blurring effect while converting a low resolution video to its high resolution counterpart. To overcome this problem, a new adaptive unsharp masking based interpolation technique is proposed here. The proposed method is a spatial domain pre-processing approach that extracts high frequency details in the spatial domain, sharpens the sub-sampled video to a certain degree through an adaptive selection of weight factor so as to compensate the blurring caused by the subsequent discrete cosine transform (DCT) based interpolation technique. The weight factor is determined adaptively for the first sub-sampled video frame by referring the corresponding original frame. Now, by exploiting the high correlation property of a video sequence, the same weight factor is utilized for sharpening the other sub-sampled video frames for restoring the fine details and critical edge information in the reconstructed video. Experimental results reveal that the proposed method shows an improvement up to 0.365 dB than DCT and a significant improvement up to 1.4 dB than the popular Bicubic interpolation technique in terms of average peak-signal-to-noise-ratio (PSNR) at 4:1 compression ratio with improved visual quality.**

Keywords-Image and video processing; Video interpolation; Unsharp masking; Discrete cosine transform (DCT); up-sampling

I. INTRODUCTION

Video frame resizing has extensively gained considerable importance in the contemporary video distribution because of its potential ability to deliver videos to various receiving devices with different display resolutions and its ability to reconstruct a high resolution video from a low resolution counterpart. This scalable feature of video interpolation not only makes the video compatible over a wide range of display devices starting from cell phone to HDTV but also results in a reduced transmission bandwidth requirement. Up-sampled high resolution videos not only give viewers a pleasant picture but also provide additional information that are necessary for subsequent analysis and interpretation in many video processing applications. In case of radar and medical systems, very often low resolution images are used because of the storage and transmission bandwidth constraints. Therefore up-sampling of the original videos is much desired in order to improve resolution for subsequent analysis such as inspection or recognition. Apart from this, scalability is one of the important features of video interpolation which plays a key role in the fast rising internet technology and consumer electronics applications. For instance browsing a remote video database, it would be more economical and convenient to send a low resolution version of the video clip to the user and then, if the user shows interest, the resolution can be enhanced using interpolation. Likewise the spatial scalable feature of video interpolation is exploited in HDTV for making it compatible with most of the existing video compression standards such as MPEG-3, MPEG-4 etc. Thus the analysis and exploitation of video interpolation are imperative for substantial improvement in the performance of the video in terms of quality, scalability and compatibility in various video processing applications.

Currently several interpolation functions are in use for image re-sampling. The simplest of these is bilinear interpolation, where the value of the new point is computed using linear interpolation of four pixels surrounding the new point [1]. More elaborate functions which consider a bigger number of pixels in interpolation process are also in use. The cubic interpolation and B-spline interpolation functions [2], [3], [4], [5] generally use sixteen pixels in the interpolation process. These commonly used methods although having simplicity and better approximation results; they suffer from some inherent defects such as blocking, blurring and ringing artifacts. Many approaches for image resizing have been developed in transformed domain. In general up-sampling in DCT domain is implemented by padding zero coefficients to the high frequency side. The various methods [6], [7] for image resizing in the transformed domain showed good result in computational complexity, scalability and image quality. However these techniques too have ringing and to some extent blurring artefacts. Thus there is a requirement of an efficient interpolation technique which provides a very less amount blurring for improved video quality.Here a hybrid and efficient video interpolation technique is proposed, which synthesizes the spatial domain adaptive unsharp masking operation with DCT domain interpolation meant for improved restored video quality. The proposed method not only provides less blurring as compared to other techniques but also gives better performance under different circumstances with different type of video data taken.

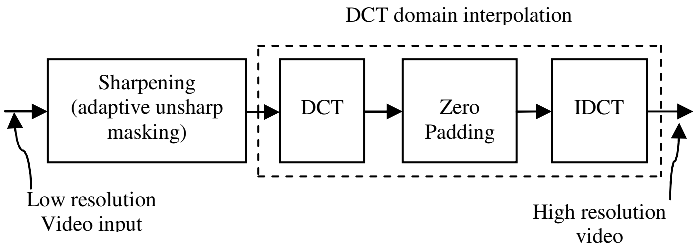


Figure 1. Adaptive unsharp masking based DCT interpolation

The proposed method is described in the subsequent section whereas experimental results are given in section-3. Finally, the work is concluded in section-4.

II. PROPOSED METHOD

In a transmitter, a sub-sampled video is produced by deleting alternate rows and columns of the original video for efficient use of a transmission channel bandwidth whereas a receiver upsamples the low resolution video to its original size by using DCT based interpolation technique. At the receiver a spatial domain pre-processing approach in the form of adaptive unsharp masking is used in order to retain some of the fine details and critical edge information in the restored video sequence.

The proposed method extracts high frequency details in the spatial domain and sharpens the sub-sampled video to a certain degree through an adaptive selection of weight factor, k so as to compensate the blurring caused by the subsequent DCT domain interpolation. The weight factor, k (described in Section-2.A) is determined adaptively for the first sub-sampled video frame by referring the corresponding original frame. Now taking the high correlation property of the video sequences into account, the same weight factor is utilized for sharpening the other sub-sampled video frames. This avoids the need for determining the weight factor for each sub-sampled video frame using the corresponding original frame. However, the weight factor is updated after a regular frame interval say after 40 or 50 frames as per the characteristics of the ongoing video sequence for more improved performance. Interpolation in either spatial or frequency domain is inadequate to achieve better performance in terms of objective and subjective quality. The proposed interpolation scheme exploits the advantages from both spatial and frequency domain processing for improved up-sampling performance. The interpolation procedure is explained in Figure 1.

A. *Adaptive Unsharp Masking*

The adaptive unsharp masking operation is performed on sub-sampled video in order to compensate the blurring, generated in the subsequent DCT domain interpolation. The degree of sharpening of the sub-sampled video sequences is performed adaptively by appropriate selection of weight factor, which is taken from the performance characteristics: PSNR vs. Weight factor plot. This characteristic plot (given in Figure 3) represents the variation of PSNR (dB) of the first restored video frame with respect to weight factor, k. The adaptive algorithm selects the weight factor corresponding to the maximum PSNR (dB) of the restored video sequence intended for the subsequent unsharp masking operation. In this method, a Gaussian mask of size (3×3) is taken instead of an averaging mask for blurring the original video. It is so because the Gaussian mask produces less blurring as compared to averaging mask and hence is suitable for improved up-sampling performance.The unsharp masking operation is used to sharpen a video frame by subtracting an unsharp or smoothed version of the video frame from the original. This operation consists of the following steps [8].

- Blur the original video frame using the Gaussian mask as indicated in Figure 2.
- Subtract the blurred video frame from the original. The resulting difference is called the mask.
- Add the mask to the original and repeat this operation for all the frames.

Let $g_1(x,y,n)$, $f(x,y,n)$ denote the blurred video sequences and original video sequences respectively. The unsharp masking is expressed in equation form as follows.

$$g_{mask}(x,y,n) = f(x,y,n) - g_1(x,y,n) \tag{1}$$

Then the weighted portion of the mask is added back to the original video frame for sharpening and is given by

$$g_2(x,y,n) = f(x,y,n) + k.g_{mask}(x,y,n) \tag{2}$$

where $g_2(x,y,n)$ and k denote sharpened video sequence and weight factor respectively. ' n ' represents the frame number that represents discrete time. The amount of video sharpening is a function of the weight factor, k . Generally for sharpening purpose, a nonzero positive weight factor is taken. In this proposed method, the value of weight factor is determined adaptively corresponding to the maximum PSNR from PSNR (dB) Vs weight factor plot of the first frame of a restored video sequence. As per the experimental evidences, the adaptive weight factor ranges in between 0 and 1 $(0 < k < 1)$ for providing better PSNR (dB) gain. It is because, for lesser value of k $(0 < k < 1)$, the sharpening will be fairly less in order to compensate the extent of blurring particularly in case of 4:1 and 16:1 compression ratio. Conversely, for $k > 1$, the sharpening will be more in comparison to blurring and this excessive sharpening will increase the mean square error of the restored video sequence.

$\frac{1}{16}\times$

1	2	1
2	4	2
1	2	1

Figure 2. A Gaussian mask used in adaptive unsharp masking operation

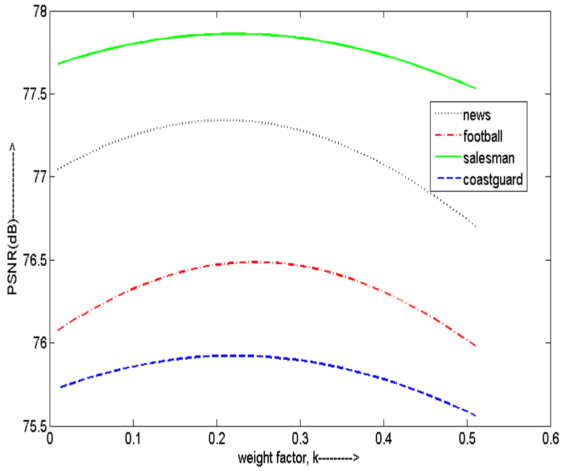


Figure 3. PSNR (dB) Vs weight factor performance characteristics of 1st frame of different videos

B. Up-sampling in DCT 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. This process was described at length in [6]. In the case of 2-D video frames, the twofold up-sampling process in a matrix form can be described as

$$b^U_{2N\times 2N} = W^T_{2N\times 2N} \times \begin{pmatrix} 2W_{N\times N} b_{N\times N} W^T_{N\times N} & 0 \\ 0 & 0 \end{pmatrix} \times W_{2N\times 2N} \qquad (3)$$

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 $N \times N$ zero matrix [9].

III. EXPERIMENTAL RESULTS

To demonstrate the performance of the proposed technique, the input video sequences are down-sampled in the spatial domain by deleting alternate rows and columns with (4:1) and (16:1) compression ratio respectively. Then for each scheme, we interpolate the frames back to their original size to allow the comparison with the original video frame. Experimental results show the proposed hybrid interpolation technique outperforms DCT and other spatial domain interpolation scheme in terms of objective and subjective video quality.

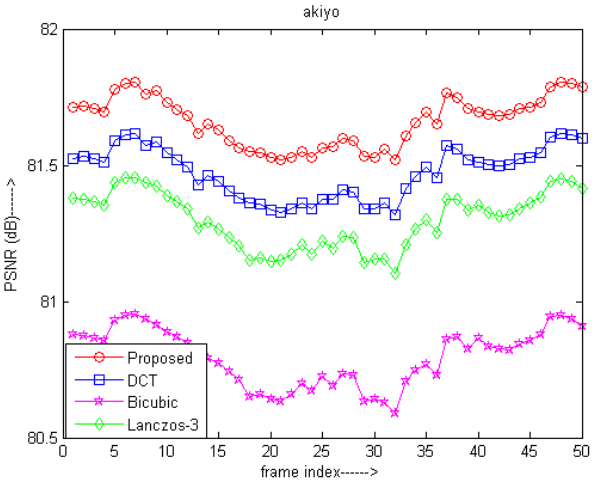
In Table I and II, we have illustrated the average PSNR (dB) comparison of various existing techniques such as DCT, bicubic, Lanczos-3, bilinear with the proposed interpolation technique in 4:1 and 16:1 compression ratios respectively. Here we have tested the proposed algorithm with various CIF, SIF and QCIF sequences and found the maximum average PSNR improvement of 0.36 dB in case of football_cif sequence at 4:1 compression ratio and 0.1dB in case of news_cif sequence at 16:1 compression ratio respectively. As per the experimental

TABLE I. AVERAGE PSNR COMPARISON OF DIFFERENT GRAY SCALE CIF(288×352), SIF(240×352) AND QCIF(144×176) SEQUENCES AT 4:1 COMPRESSION RATIO

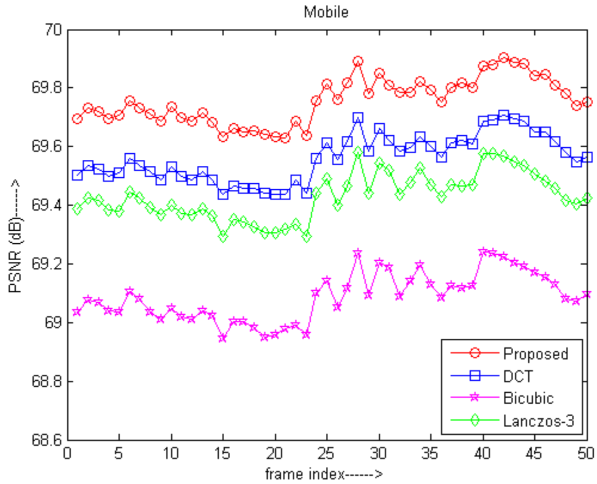
Video	**Average PSNR (dB)**				
	Proposed	***DCT***	***Bicubic***	***Lanczos3***	***Bilinear***
News_cif	77.412	77.138	76.356	76.953	74.940
Flower_sif	67.709	67.619	67.444	67.632	66.891
Football_cif	78.097	77.732	76.668	77.438	75.236
hall_cif	75.846	75.672	74.850	75.247	74.045
Mobile_cif	69.756	69.561	69.087	69.430	68.232
Salesman_cif	77.749	77.570	77.110	77.457	76.237
Stefan_sif	71.122	71.004	70.732	70.995	70.073
Foreman_cif	79.073	78.953	78.217	78.628	77.137
Coastguard_cif	75.410	75.193	74.622	75.054	73.694
Akiyo_cif	81.660	81.471	80.795	81.298	79.578
Bicycle (576×720)	78.267	77.945	76.967	77.662	75.505
City_qcif	77.051	76.930	76.176	76.476	75.514
Ice_qcif	76.778	76.636	75.547	75.893	74.844
Paris_qcif	70.483	70.373	69.979	70.190	69.398
Salesman_qcif	77.269	77.167	76.843	77.067	76.108
News_qcif	73.712	73.519	72.971	73.358	72.097

TABLE II. AVERAGE PSNR COMPARISON OF DIFFERENT GRAY SCALE CIF(288×352), SIF(240×352) AND QCIF(144×176) SEQUENCES AT 16:1 COMPRESSION RATIO

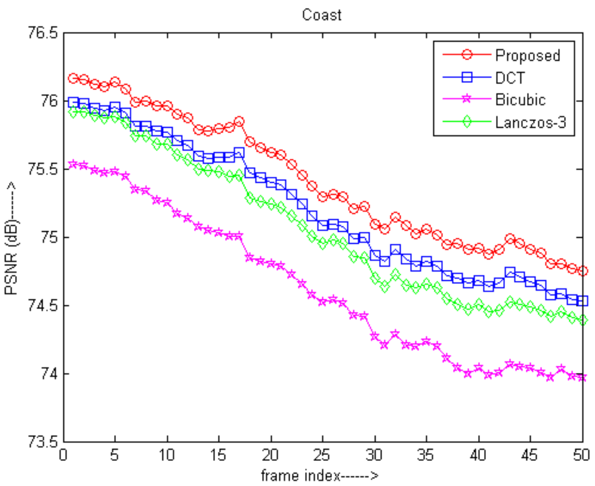
Video	**Average PSNR (dB)**				
	Proposed	***DCT***	***Bicubic***	***Lanczos3***	***Bilinear***
News_cif	70.926	70.824	70.620	70.847	70.086
Flower_sif	65.072	65.045	64.966	65.032	64.784
Football_cif	71.767	71.677	71.325	71.529	70.821
hall_cif	71.090	71.059	70.844	71.006	70.455
Mobile_cif	65.825	65.780	65.652	65.772	65.331
Salesman_cif	73.455	73.398	73.268	73.398	72.889
Stefan_sif	67.894	67.845	67.678	67.776	67.410
Foreman_cif	74.094	74.003	73.307	73.544	72.707
Coastguard_cif	70.752	70.713	70.591	70.698	70.306
Akiyo_cif	76.277	76.214	75.968	76.170	75.424
Bicycle (576×720)	72.089	71.942	71.515	71.844	70.746
City_qcif	72.936	72.912	72.639	72.777	72.325
Ice_qcif	71.974	71.940	71.508	71.715	71.079
Paris_qcif	67.207	67.177	67.090	67.165	66.782
Salesman_qcif	73.830	73.757	73.407	73.598	72.956
News_qcif	69.327	69.282	69.080	69.196	68.78



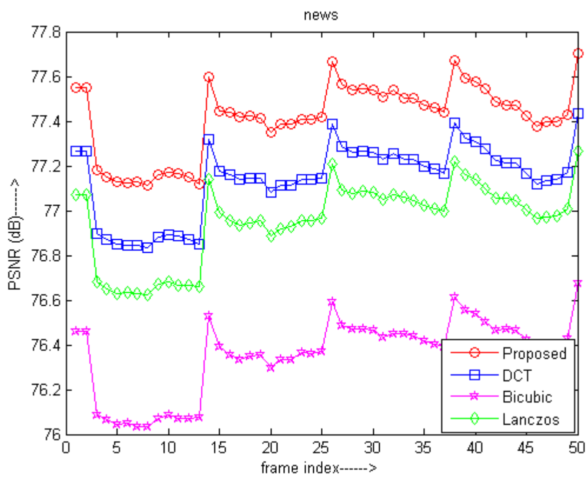
(a)



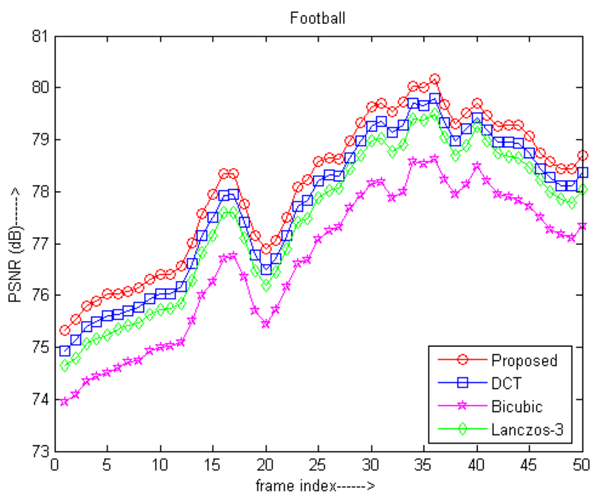
(b)



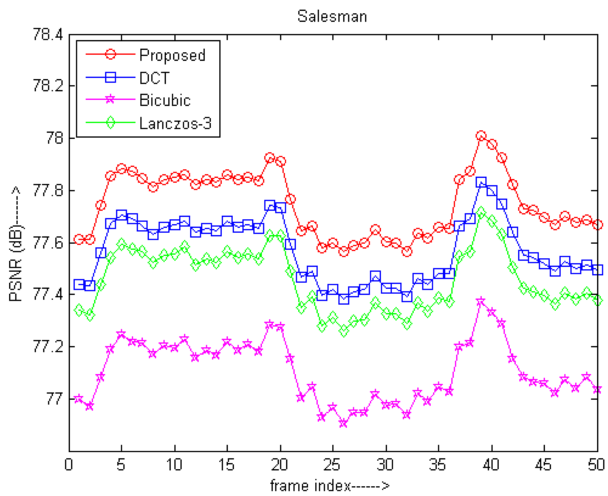
(c)



(d)



(e)



(f)

Figure 4. PSNR (dB) comparison of different sequences using different interpolation techniques at 4:1 compression ratio of:(a) akiyo; (b) mobile; (c) coastguard; (d) news; (f) salesman

results, it is quite evident that the psoposed algorithm provides a considerable PSNR improvement in comparison to other mentioned techniques irrespective of change in the resolution and compression ratio.

Figure 4 and Figure 5 represent the PSNR (dB) Vs frame index plot of different video sequences using various interpolation techniques at 4:1 and 16:1 compression ratio respectively. Experimental results show, the proposed method yields better objective quality for different types of sequences having dissimilar characteristics.

In Figure 6, Figure 7 and Figure 8 we have illustrated the results of various up-sampling schemes at 4:1 compression ratio for subjective evaluation. Experimental results show, the blurring is much reduced and the edges are more pronounced with fine detail preservation in comparison to other existing interpolation techniques. Figure 6 and Figure 7 represent the subjective performance of the 40th frame of coastguard and salesman sequence respectively. Figure 8 represents the subjective performance of the 11th frame of football sequence. Thus it is quite obvious that the proposed method yields considerably better subjective performance than DCT and other techniques irrespective of the variation in video characteristics.

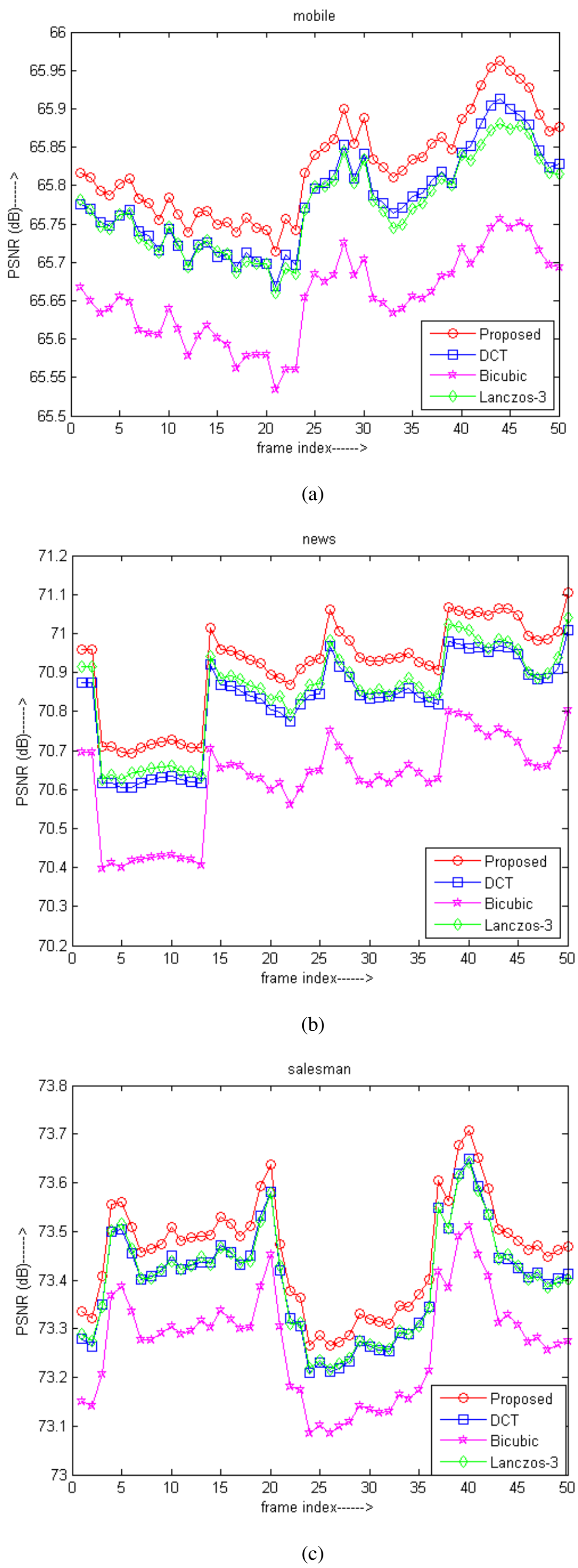


Figure 5. PSNR (dB) comparison of different sequences using different interpolation techniques at 16:1 compression ratio of:(a) mobile; (b) news; (c) salesman.

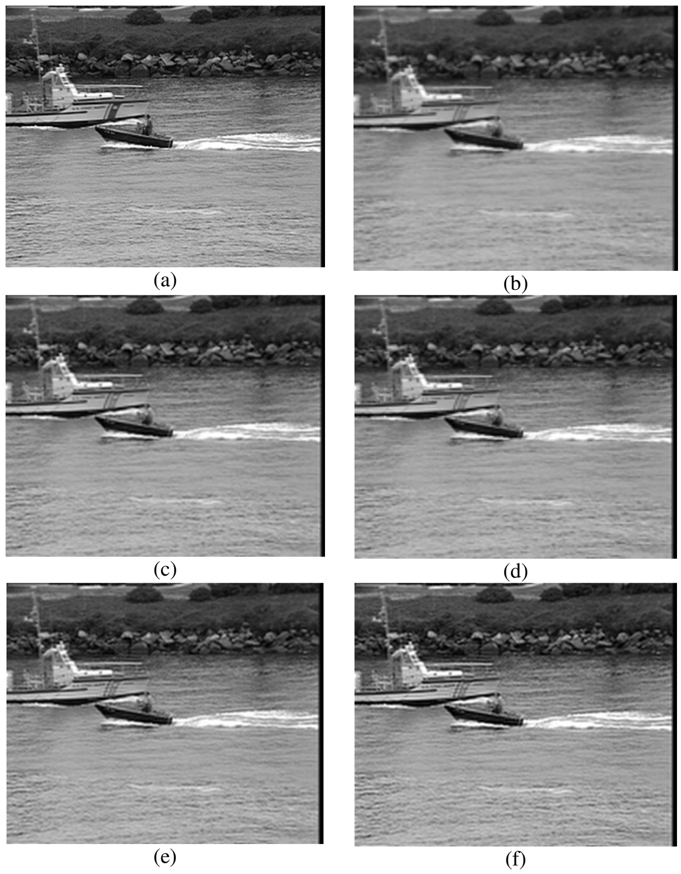


Figure 6. Subjective performance of the 40th frame of the football sequence at 4:1 compression ratio using different interpolation techniques: (a) original; (b) bilinear; (c) Lanczos-3; (d) bicubic; (e) DCT; (f) proposed.

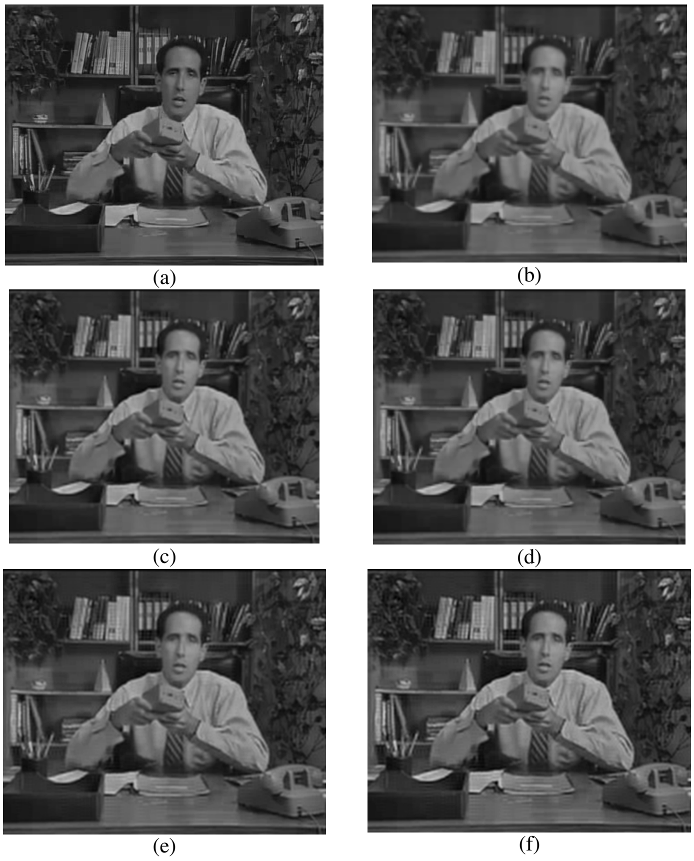


Figure 7. Subjective performance of the 40th frame of the salesman sequence at 4:1 compression ratio using different interpolation techniques: (a) original; (b) bilinear; (c) Lanczos-3; (d) bicubic; (e) DCT; (f) proposed.

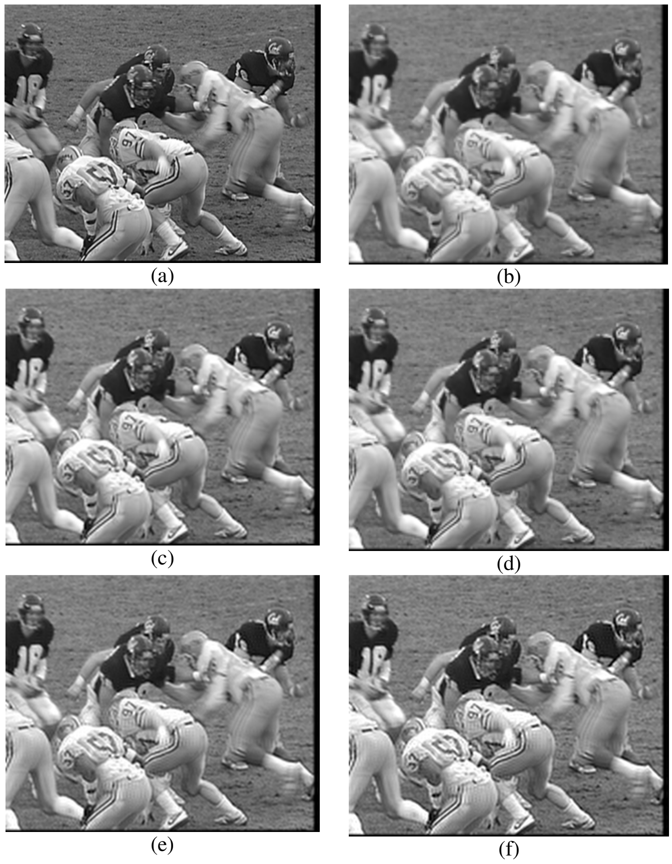


Figure 8. Subjective performance of the 11th frame of the football sequence at 4:1 compression ratio using different interpolation techniques: (a) original; (b) bilinear; (c) Lanczos-3; (d) bicubic; (e) DCT; (f) proposed.

IV. CONCLUSION

Here a new interpolation technique is proposed which not only restores a sub-sampled video with high precision but also yields a very low degree of blurring with fine detail preservation through an adaptive selection of weight factor, k. The proposed algorithm exploits the advantages of both spatial domain and frequency domain processing for improved up-sampling performance in terms of PSNR (dB). It delivers superior performance and high degree of flexibility under a variety of constraints such as variation in resolution, compression ratio and video characteristics. It works better preferably at high resolution and low compression ratio but at the same time is flexible enough to provide considerable performance at low resolution and high compression conditions. In addition, with the adaptive unsharp masking operation, it works fine with different types of videos having dissimilar characteristics and thus achieves better subjective performance. Since the proposed method is a preprocessing approach, it imparts more computational burden on the transmitting side than the receiving end and thus makes the receiver computationally less complex, fast and suitable for various real time applications. Moreover, the exploitation of high correlation property of a video sequence in the proposed algorithm makes it faster, less complex, more efficient and suitable for different up-sampling applications.

REFERENCES

[1] 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.

[2] 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.

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

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

[5] H. S. Hou and H. C. Andrews, “Cubic splines for image interpolation amd digital filtering” IEEE Trans. Acoust., speech and sign. Proc., vol. ASSP-26, 1978.

[6] 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.

[7] J. Mukherjee and S. K. Mitra, “Image resizing in the compressed domain using subband DCT,” IEEE Trans. Circuits, Syst., Video Technology, vol. 12, pp. 620-627, July 2002.

[8] Rafael Gonzalez and Richard Woods, Digital Image Processing, Pearson Publications.

[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.