Region Adaptive Unsharp Masking based DCT Interpolation for Efficient Video Intra Frame Up-sampling

Aditya Acharya
Dept. of Electronics and Communication Engg.
National Institute of Technology
Rourkela-769008, Odisha, India

Sukadev Meher
Dept. of Electronics and Communication Engg
National Institute of Technology
Rourkela-769008, Odisha, India

ABSTRACT

Most of the existing interpolation techniques available in the literature produce a blurring effect while converting a low resolution video to its high resolution counterpart. The blurring results in the loss of fine details and critical edge information of a video intra frame. In order to resolve this problem, an efficient, no reference, hybrid interpolation technique is proposed here. The proposed method makes use of a combination of anticipatory, spatial domain, region adaptive unsharp masking operation with the discrete cosine transform (DCT) based interpolation technique for retaining some of the fine details and critical edge information in the reconstructed video frame. The region adaptive unsharp masking is a preprocessing step which sharpens the intra frame regions locally as per their statistical local variance so as to compensate the blurring caused by the subsequent DCT based interpolation technique. The degree of sharpening is proportionately increased if the local variance is greater than the global variance on the contrary the sharpening will be proportionately reduced, if the local variance is less than the global threshold value. Experimental results reveal that the proposed method outperforms most of the existing interpolation techniques in terms of peak-signal-to-noise-ratio (PSNR) as well as visual quality for different types of video sequences.

General Terms

Video restoration, video resizing, interpolation.

Keywords

Image and video processing, Video interpolation, Unsharp masking, Discrete cosine transform (DCT), Up-sampling

1. INTRODUCTION

Video frame resizing has gained much importance in the contemporary video communication because of its potential features like scalability and compatibility with various receiving devices with different display resolutions. This scalable feature is because of the interpolation technique which makes the video compatible over a wide range of display devices starting from mobile phones to HDTV. Frame resizing also plays a key role in reducing the transmission bandwidth requirement which consequently avoids channel congestion.

Up-sampled high resolution video not only gives a better HVS performance to a viewer but also provides additional information for various post processing applications such as inspection or recognition. In medical imaging, remote sensing and video surveillance applications, very often it is desired to improve the native resolution offered by imaging hardware for subsequent analysis and interpretation. Video interpolation aims to generate high resolution video from the associated low resolution capture and hence is very essential for the aforesaid applications.

Scalability is one of the key features of video interpolation which is exploited in internet technology and consumer electronics applications. For instance, while remote browsing a video database, it would be more convenient and economical to send a low resolution version of a video clip to the user. If the user shows interest the resolution can be progressively enhanced using interpolation. Similarly HDTV exploits the scalable feature of video interpolation for its compatibility with most of the existing video compression standards such as H.263 and H.264. In addition, the video is made adaptive to variable bit rate and computational capacities of different receiving devices by utilizing the scalable feature of interpolation. Thus the analysis and exploitation of video interpolation are quite essential to improve the performance of contemporary video communication in terms of quality, scalability and compatibility.

Currently several interpolation techniques are used in video resampling process. One of the simplest interpolation technique is a nearest neighbor interpolation. In this case, the value of a new point in the interpolated image is taken as the value of old coordinate which is located nearest to the new point. Although it is a simple technique, it suffers through blocking artifacts. Another frequently used technique is bilinear interpolation where the value of a new point is computed using linear interpolation of four pixels surrounding the new point [1]. Bilinear interpolation though is simple and less complex, it has undesirable blurring artifacts. There are widely used interpolation techniques such as bicubic and B-spline [2-5] which consider sixteen pixels for determining a new interpolated point. These techniques provide better performance in terms of quality at the cost of computational complexities. Bicibic and B-spline interpolation techniques provide a less degree of blurring in comparison to bilinear 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 range [6], [7]. Even if Lanczos interpolation gives good results, it is slower than other approaches and provides a blurring effect in the reconstructed image. Many approaches for image resizing have been developed in transform domain. Up-sampling in DCT domain is implemented by padding zero coefficient to the high frequency side. Image resizing in DCT domain shows very good result in terms of scalability and image quality. However, this technique suffers through undesirable blurring and ringing artifacts. Thus there is a requirement of efficient

interpolation technique which not only produces a very least degree of blurring and ringing but also provides improved objective performance in the reconstructed video.

The organization of the paper is structured as follows. The proposed method is described in section-2. Section-3 provides the simulation results of different interpolation algorithms subjected to various constraints. Finally the work is concluded in section-4.

2. PROPOSED METHOD

Generally in a transmitter, a sub-sampled video is produced by alternate deletion of rows and columns for effective use of transmission channel bandwidth where as at the receiver, the resolution of the sub-sampled video is enhanced using the suitable interpolation technique. The proposed method is an anticipatory, spatial domain, preprocessing operation which is coupled with the DCT domain interpolation scheme in order to retain some of the fine details and critical edge information that is lost during the sub-sampling operation and while converting a low resolution video to it's high resolution counterpart.

A DCT based up-sampling scheme has an important property of preserving the low frequency components generated by smooth and fast changing area in a video intra frame. This consequently results in the loss of high frequency details which leads to blurring in the up-sampled video intra frame. To overcome this problem, we make use of a sharpening technique in spatial domain to compensate the high frequency loss. Therefore, a hybrid interpolation technique is proposed here which exploits the advantages of transform domain DCT based interpolation and spatial domain region adaptive unsharp masking for preserving the fine details and critical edge information in the up-sampled video intra frame. Here the high frequency loss due to DCT based interpolation is compensated by a preprocessing spatial domain sharpening technique which makes use of region adaptive unsharp masking operation.

The region adaptive unsharp masking operation is a preprocessing step that sharpens the sub-sampled video intra frame to a certain degree depending on it's statistical local variance. The local regions with high local variance are proportionately sharpened more than the regions with less local variance by the proposed adaptive algorithm. In order to perform this operation, the global variance of a video intra frame is computed as global threshold. The statistical local variance of a region is compared with the global threshold value. The degree of sharpening is proportionately increased if the local variance is greater than the global threshold value on the contrary the sharpening will be proportionately reduced, if the local variance is less than the global threshold value. As a consequence of this, the fast changing high frequency regions are sharpened more than the slowly varying smooth regions so as to compensate the high frequency loss. Since the DCT based interpolation scheme preserves the low frequency components quite effectively, the degree of sharpening has been made proportionately less in case of slowly varying smooth regions. The proposed method consists of basic two steps namely region adaptive unsharp masking and DCT domain interpolation and are described subsequent subsections.
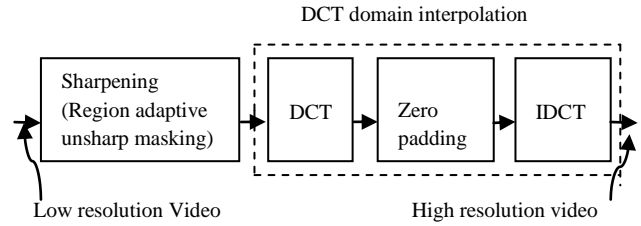

Fig 1: Region adaptive unsharp masking based DCT interpolation technique

2.1 Region Adaptive Unsharp Masking

The region adaptive unsharp masking process is used to sharpen a video frame by subtracting an unsharp or smoothed version of it from the original. The smooth version of a video frame is obtained by blurring it using a region adaptive Gaussian mask whose center pixel weight is varied depending on the statistical local variance of a neighborhood. The region adaptive algorithm computes the global variance of a video intra frame as a threshold. The statistical local variance of a neighborhood is then compared with this global threshold value. If local variance of the neighborhood is larger than the global threshold value, the weight of the center pixel is reduced proportionately in order to provide more blurring to the high variance region. This consequently results in more degree of sharpening in the regions having high variance. Similarly, the reverse operation is performed in case of the low variance regions. Subsequently, the blurred video frame is subtracted from the original to form an unsharp mask. This mask is then added to the original video frame to form a sharpened video intra frame. In brief, the region adaptive unsharp masking consists of the following steps [8].

- Blur the original video frame using a region adaptive Gaussian mask.

- 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 by region adaptive Gaussian mask and original video sequence respectively. The region adaptive unsharp masking is expressed in equation form as follows

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

$g_1(x, y, n)$ is obtained by using the region adaptive blurring algorithm. Then the mask is added back to the original video frame for sharpening and is given by

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

Where $g_2(x, y, n)$ denotes the sharpened video sequence. n represents the frame number that represents discrete time. The region adaptive Gaussian mask for sharpening operation is given in Figure 2. W_t represents the center pixel weight of the mask which is made adaptive as per the statistical local variance of a 3×3 neighborhood.

Region adaptive algorithm for determining $g_1(x,y,n)$

---

for $n = 1$ to frame number do

Find global mean M and global variance V for each frame

$$M \leftarrow \frac{1}{PQ}\sum_{x=1}^{P}\sum_{y=1}^{Q} f(x,y,n)$$

$$V \leftarrow \frac{1}{PQ}\sum_{x=1}^{P}\sum_{y=1}^{Q}[f(x,y,n)-M]^2$$

for $x = 1$ *to* P

for $y = 1$ *to* Q

Find local mean m and local variance v for each

Pixel in a neighborhood

$$m \leftarrow \frac{1}{9}\sum_{s=-1}^{1}\sum_{t=-1}^{1} w(x+s,y+t)$$

$$v \leftarrow \frac{1}{9}\sum_{s=-1}^{1}\sum_{t=-1}^{1}[f(x+s,y+t)-m]^2$$

if $v > 1.5 \times V$

$w_t = 2$

elseif $v > V$ *and* $v \le 1.5 \times V$

$w_t = 4$

elseif $v > (3V/4)$ *and* $v \le V$

$w_t = 8$

elseif $v > (V/2)$ *and* $v \le (3V/4)$

$w_t = 16$

elseif $v > (V/4)$ *and* $v \le (V/2)$

$w_t = 32$

else

$w_t = 64$

end

$$h \leftarrow \frac{1}{w_t+12} \times \begin{bmatrix} 1 & 2 & 1 \\ 2 & w_t & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

$$g_1(x,y) \leftarrow \frac{1}{w_t+12}\sum_{s=-1}^{1}\sum_{t=-1}^{1} h(s,t)f(x+s,y+t)$$

end

end

end

$$\frac{1}{12+w_t} \times \begin{array}{|c|c|c|} \hline 1 & 2 & 1 \\ \hline 2 & w_t & 2 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

Fig 2: Region adaptive Gaussian mask

2.2 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. Subsequently, 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 intra frames, the twofold up-sampling process in a matrix form can be described as

$$b_{2N \times 2N}^U = W_{2N \times 2N}^T \times \begin{pmatrix} 2W_{N \times N}b_{N \times N}W_{N \times N}^T & 0 \\ 0 & 0 \end{pmatrix} \times W_{2N \times 2N} \quad (3)$$

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

3. EXPERIMENTAL RESULTS

To demonstrate the performance of the proposed hybrid technique, the input video sequences are down-sampled in the spatial domain by deleting alternate rows and columns at (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.

Table 1 and Table 2 illustrate the average PSNR comparison of DCT, bicubic, lanczos-3 and the proposed interpolation techniques at 4:1 and 16:1 compression ratios respectively. Experimental results reveal that the proposed technique shows up to 0.4 dB average PSNR gain than DCT and a gain up to 1.5 dB than the popular Bicubic interpolation technique at 4:1 compression ratio particularly in the case of football sequence. Similarly the proposed technique achieves a gain up to 0.1dB than DCT at 16:1 compression ratio in case of news sequence. The PSNR gain at 4:1 compression ratio is more than the PSNR gain at 16:1 compression ratio. It is because, at a high compression ratio, most of the high frequency details are lost, finally giving a flat and blurred output. Since the proposed method employs the high frequency details of the subsampled video for sharpening it for original video restoration, the PSNR gain at high compression ratio is less than the low compression counterpart.

In Figure 3, we have shown the variation of PSNR with respect to the frame index at 4:1 compression ratio. In Figure 4, the subjective performances of different interpolation techniques are illustrated for the 11th frame of different sequences at 4:1 compression ratio. 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 irrespective of the video types.

Table 1. Average PSNR (dB) of different CIF and QCIF video sequences at 4:1 compression ratio
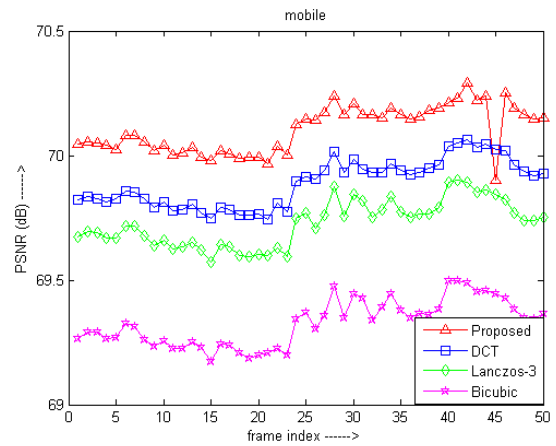
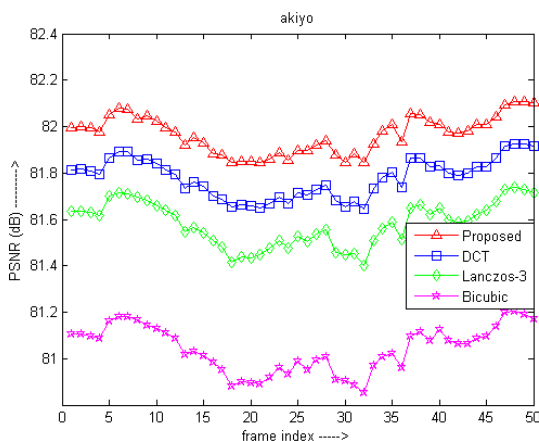| Videos	Average PSNR (dB)			
	Bicubic	**Lanczos3**	**DCT**	**Proposed**
Akiyo_cif	81.0421	81.5810	81.7782	**81.9659**
News_cif	76.4648	77.0882	77.2933	**77.5837**
Salesman_cif	77.1100	77.4577	77.5707	**77.7493**
Football_cif	76.6988	77.4973	77.8178	**78.2245**
Soccer_cif	78.3849	78.7956	78.8791	**78.9670**
Hallmonitor_cif	74.5728	74.9829	75.4166	**75.6101**
Mobile_cif	69.3281	69.7275	69.8889	**70.1052**
Bus_cif	72.1562	72.6002	72.6701	**72.8476**
Akiyo_qcif	76.4515	76.7693	77.4747	**77.6849**
Football_qcif	73.3062	73.5380	73.7264	**73.8292**
Signirene_qcif	80.2644	80.5730	80.7755	**80.8635**
Hallmonitor_qcif	73.3966	73.7401	74.0960	**74.1849**
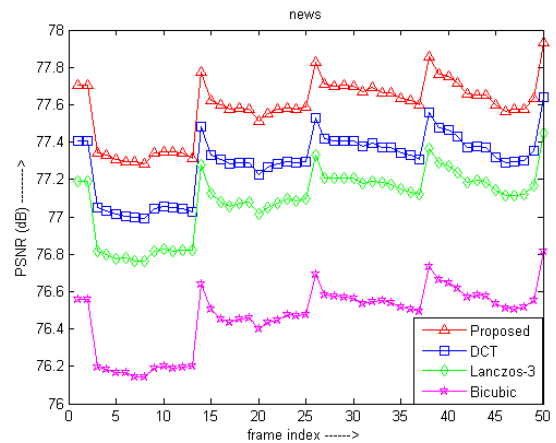
**Table 2. Average PSNR (dB) of different CIF and QCIF video sequences at 16:1 compression ratio**

Videos	Average PSNR (dB)			
	Bicubic	**Lanczos3**	**DCT**	**Proposed**
Akiyo_cif	76.2144	76.4112	76.4478	**76.4902**
News_cif	70.7512	70.9707	70.9456	**71.0372**
Salesman_cif	73.2679	73.3983	73.3981	**73.4530**
Football_cif	71.3216	71.5169	**71.6358**	70.4207
Soccer_cif	74.2612	74.4484	74.4829	**74.4845**
Hallmonitor_cif	70.6006	70.7611	70.8146	**70.8297**
Mobile_cif	65.8343	65.9303	65.9213	**65.9566**
Bus_cif	69.0830	69.1872	69.1570	**69.1737**
Akiyo_qcif	72.9350	73.0811	73.3088	**73.3286**
Football_qcif	70.3124	70.4294	70.5052	**70.5190**
Signirene_qcif	76.6243	76.8263	76.8986	**76.9259**
Hallmonitor_qcif	69.7317	69.8765	70.0648	**70.0683**

(b)

(c)

(a)
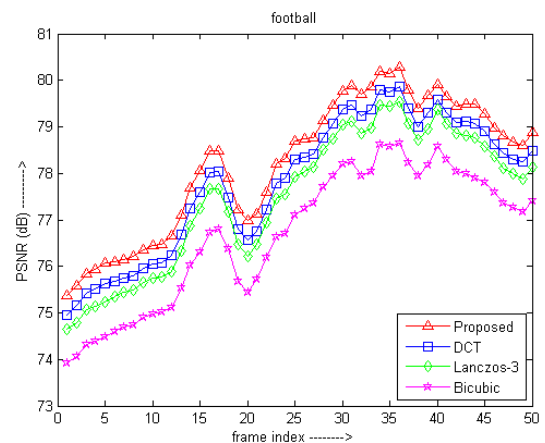
(d)

Fig 3: Average PSNR (dB) comparison of different video sequences using various interpolation technique at 4:1 compression ratio : (a) akiyo_cif; (b) new_cif; (c) mobile_cif; (d) football_cif.

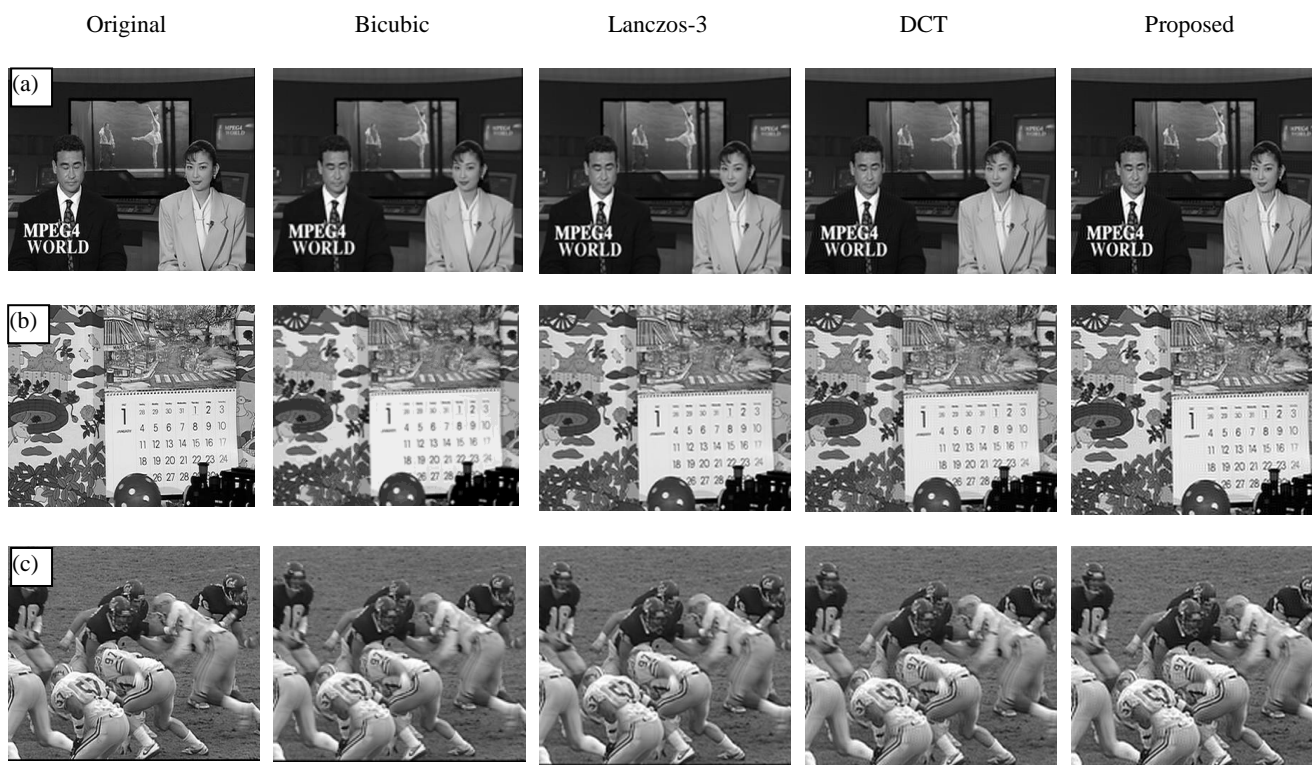| Original	Bicubic	Lanczos-3	DCT	Proposed

Fig 4: Subjective performance of the 11th frame of different video sequences at 4:1 compression ratio using different interpolation techniques: (a) news_cif; (b) mobile_cif; (c) football_cif.

4. CONCLUSION

Here a no reference hybrid 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 by exploiting the advantages of both spatial domain and frequency domain technique. It delivers superior performance and high degree of flexibility under a variety of constraints such as variation in resolution, compression ratio and the type of a video sequence. It works preferably better at high resolution and low compression ratio but at the same time is flexible enough to provide considerable performance under low resolution and high compression conditions. In addition, by making use of region adaptive unsharp masking operation, it works fine with different types of videos having dissimilar characteristics and thus achieves better subjective and objective performance. Since the proposed method is a preprocessing operation, 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. Thus the proposed method is a low complex, highly flexible and efficient algorithm that works fine with different types of video data.

5. REFERENCE

[1] Jing, L., Si, X., and Shihong, W. 2009. An improved bilinear interpolation algorithm of converting standard definition to high definition images. In proceedings of the WASE Int. Conf. On Info. Engg., 441-444.

[2] Keys, R. G. 1981. Cubic convolution interpolation for digital image processing. IEEE. Trans. Acoust. Speech. Signal Process. ASSP-29 (Dec. 1981), No. 6, 1153-1160.

[3] Reichenbach, S. E. and Geng. F. 2003. Two-dimensional cubic convolution. IEEE. Trans. Image Process. 12 (Aug. 2003), 857-865.

[4] Dengwen, Z. 2010. An edge directed bicubic interpolation algorithm. CISP, 1186-1189.

[5] Hou, H. S. and Andrews H. C. 1978. Cubic spline for image interpolation and digital filtering. IEEE. Trans. Acoust. Speech and Sign. Process. ASSP-26.

[6] Dugad, R. and Ahuja, N. 2001. A fast scheme for image size change in the compressed domain. IEEE. Trans. Circuit Syst. Video Tech. 11 (Apr. 2001), 461-474.

[7] Mukherjee, J. and Mitra, S. K. 2002. Image resizing in the compressed domain using subband DCT. IEEE. Trans. Circuit Syst. Video Tech. 12(July 2002), 620-627.

[8] Gonzalez, R. and Woods, R. 2009. Digital image processing. Pearson publications.

[9] Wu, Z., Yu, H., and Chen, C. W. 2010. A new hybrid DCT based interpolation scheme for video intraframe up-sampling. IEEE signal processing letters. 17 (oct. 2010), 827-830.