MULTI-DIRECTIONAL INTERPOLATION FOR SPATIAL ERROR CONCEALMENT

Wilson Kwok and Huifang Sun
David Sarnoff Research Center, Princeton, NJ 08543

Abstract

This paper presents an algorithm for lost signal restoration in block-based still image and video sequence coding. The error concealment algorithm presented herein is aimed at masking the effect of missing blocks by using spatial interpolation. This interpolation algorithm utilizes spatially correlated edge information from a large local neighborhood of surrounding pels and performs multi-directional interpolation to restore the missing block. Results show that this algorithm is capable of restoring a block containing multiple edges that are consistent with edges present in neighboring pel regions.

1. Introduction

In block transform coding, the spatial redundancies within blocks are removed and the energy is compacted into a small number of coefficients after the transformation. Compression is achieved by quantizing the high energy coefficients with finer resolution than low energy coefficients, and in conjunction with variable length coding such as run-length coding and Huffman coding. In many tests, compression techniques based on the most prevalent block transform coding method, the discrete cosine transform (DCT), can provide good quality reconstructed images with higher than 20 to 1 compression ratio [1]. The compressed data can be stored or be transmitted through a communication channel. Practical communication channels are not error free, although the loss mechanism may vary widely from media to media. Data corruption may be caused by network congestion, thermal noise, switch noise, signal fade, etc. Since the signals transmitted on the real-world channels are highly compressed, independent of cause, the quality of images reconstructed from any corrupted data can be very unsatisfactory.

The problem of lost data in block-coded images due to imperfect communication channels needs to be solved. Error concealment is intended to ameliorate the impact of channel impairments (i.e. bit-errors in noisy channels, or cell-loss in packet networks) by utilizing a-priori information about typical images in conjunction with available picture redundancy to provide subjectively acceptable restoration of damaged picture regions. In particular, a good spatial interpolation method is necessary for hiding the effect of missing blocks in still images and video frames. Temporal interpolation, or replenishment, by itself is not always adequate for concealment of errors in video sequences. This is especially true for stressful image sequences with irregular motion, abrupt scene changes, and intra-coded image frames which are treated as still images. To the viewer, poor temporal replacement of error regions appear as portions of video being broken up into displaced pieces. Earlier consideration of this problem was addressed in [2,3]. In those works, the spatial error concealment used an approach of interpolating a single-pel wide boundary around the missing block to achieve a restoration based on an optimal measure of smoothness. The damaged blocks were restored fairly well in very low frequency portions of the image. However, the smoothing process restores blurry blocks with a significant loss of detail in higher frequency portions of the image.

Thereafter, we sought to utilize spatially correlated information more thoroughly by performing interpolation based on a large local neighborhood of surrounding pels and to restore edges which are continuous with those present in the neighborhood. Edge integrity plays an important role in visual perception [7]. Often, subjective viewer evaluation of images that are sub-optimally restored with emphasis on object edges are preferred over optimally smooth minimum mean-square error (MMSE) restorations. In [4], the approach used for spatial error concealment was based on projections onto convex sets (POCS). That method used a directional constraint imposed on the frequency domain description of the pel region consisting of the missing block and its surrounding neighborhood. The other constraint was imposed in the spatial domain which required surrounding pels to be left unchanged. The directional constraint was determined by an edge classifier which picks out the most likely orientation of edges in the surrounding neighborhood. Good results were obtained when the missing block can be characterized by a single dominant edge direction.

In this paper, we attempt to restore more detail in the missing block when the surrounding pels strongly suggest that more than one edge direction should be used. We take a somewhat different approach to this multi-directional edge restoration problem than the method of POCS. Since the iterative process required by POCS is computationally expensive and difficult to implement in hardware, we opt to perform directional interpolation directly in the spatial pel domain. No iterations are required by this process to restore sharp edges. Experimental results show that detailed edges can be restored in cases where the surrounding pels contain highly correlated edge information. The rest of this paper is organized as follows. Background on video bitstream packetization and transport is discussed in Section 2. In Section 3, the restoration process is presented. Section 4 provides the experimental results and discussion. Finally, concluding remarks are given in Section 5.

2. Packetization and Transport Assumptions

A typical block-based video source coder, such as MPEG [1], consists of the cascade of a linear transform operation, quantization, and entropy coding. Specifically in the MPEG standard, an image is segmented into non-

Manuscript received June 11, 1993

overlapping blocks; then each block or prediction residual block is transformed via DCT to remove spatial correlation and subsequently the DCT coefficients are quantized and entropy coded using variable length codewords. When bit errors occur in such a highly compressed bitstream, all subsequent information becomes useless until bitstream synchronization can be re-established. Packetization is the most common way to localize errors in a bitstream, and provides for resynchroniztion in the case of bit errors. Packetized bitstreams are suitable for transmission via a broadcast RF channel or packet switched network. In either scenario, bit errors that occur may lead to lost packets. In the packet network context, network congestion may cause some packets to be discarded and simply not sent to the receiver. In the broadcast RF context, damaged packets may be received with uncorrectable bit errors, and there is no way to ascertain how much of the data within the packet is usable. So for practical purposes, damaged packets are treated as lost packets.

Packets contain a known number of image data segments. For our purposes, a basic unit of image data will be taken as a 16 x 16 pixel block of image samples. A packet will contain one or more of these blocks. Packet losses therefore manifest themselves as an erasure of block pel regions in video images. It is assumed that an appropriate transport mechanism exists which can detect erroneous/lost packets, precisely locate damaged pel regions, and resume normal decoding (See for example, [5]). The shape of the damaged pel region due to a packet loss depends on how the coded blocks are grouped together for packetization. Packets may contain bits coded from non-adjacent image blocks following a staggered checkerboard scanning order or packets may comprise of bits coded from adjacent image blocks following a raster-scan order as in MPEG. In the following, we first consider the restoration of isolated lost blocks surrrounded by good blocks. Later, we address situations where adjacent blocks are also lost.

3. The Restoration Process

Restoration of lost pel values in an image is a surface fitting problem: given a partial set of surface height measurements, the task is to generate a complete surface representation. This problem is ill-posed, and as such, there is no unique solution. Additional constraints known a-priori must be used to restrict the otherwise infinite number of possible solutions. There are a-priori properties about typical video images we would like to use in performing the interpolation; these include:

1) Smoothness - requires reconstructed samples to be smoothly connected with adjacent samples
2) Edge Continuity - requires that edges of objects in the scene be continuous
3) Consistency with known values - requires that correctly received sample values not be altered by the restoration process, and that restored values lie in a known range (e.g. [0-255])

A formal treatment of lost image block restoration based on the theory of projections onto convex sets is presented in [4]. Here, we present a hueristic approach which is more intuitive and computationally cheaper to perform. A block diagram illustrating the general principle of the restoration process is shown in Figure 1. The missing block is surrounded by a neighborhood of good pels. The classifier operates on the surrounding neighborhood and determines which directions characterize the strongest edges. For each of these classified directions, spatial directional interpolation is performed to create a set of blocks, each with strong edges in their respective directions. The blocks in this set are subsequently mixed together in such a way that all the strong features of each block are extracted and combined together into a single block. Thus, we restore a detailed block containing multiple interpolated edges.

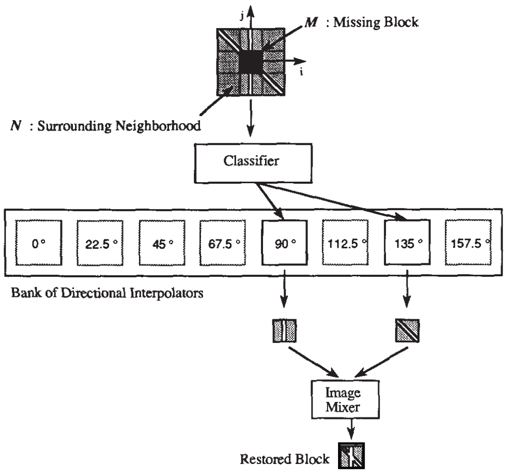


Figure 1. The multi-directional restoration process.

3.1 Directional Classification by Gradient Voting. Lost image blocks are restored by extending edges present in the surrounding neighborhood so that they pass through the missing block. To accomplish this, the most likely edge orientations should be correctly chosen based on some knowledge of the edge characteristics around the missing block. A reasonably simple and effective method of performing this classification is through the use of gradient measures in the spatial domain. The local edge gradient components for the pixel $x(i,j)$ is computed by:

$$g_x = x_{i+1,j-1} - x_{i-1,j-1} + 2x_{i+1,j} - 2x_{i-1,j} + x_{i+1,j+1} - x_{i-1,j+1} \quad (1)$$

$$g_y = x_{i-1,j+1} - x_{i-1,j-1} + 2x_{i,j+1} - 2x_{i,j-1} + x_{i+1,j+1} - x_{i+1,j-1} \quad (2)$$

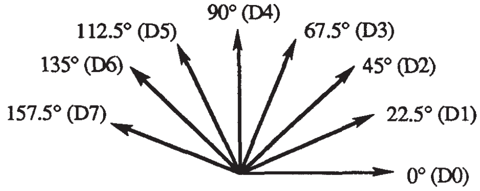


Figure 2. Eight directional edge categories.

This is equivalent to applying the following 3 x 3 Sobel mask operators:

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{3}$$

The magnitude and angular direction of the gradient at coordinate *(i,j)* are:

$$G = \sqrt{g_x^2 + g_y^2} \qquad \theta = \tan^{-1}(g_y / g_x) \tag{4}$$

The Sobel operator is selected for gradient estimation due to its circularity property [6], which gives more accurate angle estimates over the standard gradient operator. Figure 1 illustrates the missing block of pixels, denoted by M, surrounded by a neighborhood of correctly received pixels, denoted by N. The gradient measure is computed for every *(i,j)* coordinate in the neighborhood surrounding the missing macroblock. The value of the gradient angle is rounded to the nearest 22.5° and thus corresponds to one of eight directional categories equally spaced around 180°, as depicted in Figure 2. There are counters for each of the eight directions, D_0 through D_7. A voting mechanism is used which involves incrementing the selected category counter by the magnitude of the gradient if a line drawn through the pixel at (i,j) with orientation θ passes through the missing block. This is described by the following pseudo-code:

```
DO [ over all (i,j) pel coordinates in $N$ ] {
    Compute $G$ and $\theta$ from equation (4)
    k=[round($\theta$ / 22.5°) + 8 ] mod 8
    if [ line drawn through (i,j) with
          angle $\theta$ intersects $M$ ] {
          $D_k = D_k + G$
    }
}                                                    (5)
```

After all the pels in the surrounding neighborhood have "voted," the top one, two or three counters containing the largest values within a certain threshhold, T, of the maximum counter value determines the set of directions, S, to use in the interpolation:

```
$D_{\max} = \max_k (D_k)$
DO [ k = 0 to 7 ] {
    if [ $D_k > (1 - T) D_{\max}$ ] $D_k \in S$
}
if [ size($S$ ) > 3 elements ] {
    keep 3 largest $D_k$'s and remove the rest
}                                                    (6)
```

In practice, a threshhold value of $T = 0.25$ works well for a large class of natural image blocks.

3.2 Spatial Interpolation. Spatial interpolation is performed for each of the directions determined by the classifier. For a given direction, a series of one-dimensional interpolations are carried out along that direction. Figure 3 shows this process. In each of the one-dimensional interpolations, pel values are gathered along a directed line cutting through the surrounding neighborhood. We obtain a one-dimensional array of values, $x(i)$, which resembles Figure 4. All of the missing pels, $x(i) : i \in M$, are interpolated from a weighted average of neighborhood pels. The weights depend inversely with the distance, d, from the missing pel to the good neighborhood pel. Missing pels are thus interpolated as follows:

$$x_{i\,:\,i \in M} = \frac{\dfrac{x_0}{d_{i0}^w} + \dfrac{x_1}{d_{i1}^w} + \ldots + \dfrac{x_N}{d_{iN}^w}}{\displaystyle\sum_{j \notin M} \frac{1}{d_{ij}^w}} \tag{7}$$

where d_{mn} is the distance from $x(m)$ to $x(n)$ and w specifies the power of the inverse weighting. A power weighting factor of $w = 2.5$ works well in practice. The interpolated values of $x(i)$ computed from (7) are copied back into corresponding pel positions of the missing block.

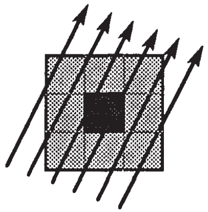


Figure 3. Series of one-dimensional interpolations.

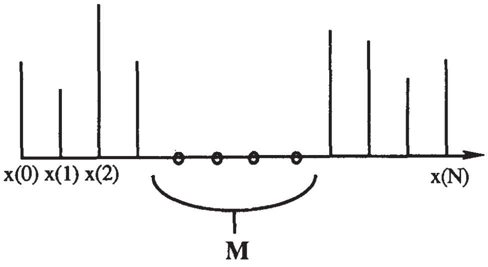


Figure 4. One-dimensional interpolation of missing pels.

3.3 Image Mixing. The purpose of image mixing is to extract the strong characteristic features of two or more images and merge them into one image. To understand how this is accomplished, one must characterize what is considered *foreground* in an image and what is considered *background*. Consider the case of mixing two images. This may be done by computing the histogram of each image, along with their means and variances. Figure 5 shows the procedure. Pels with values close to the mean may be considered background pels while pels deviating from the mean are foreground pels. A threshhold of one standard deviation works well in practice. Pels that exceed the threshhold value away from the mean and are greater than the mean correspond to bright foreground features and those that exceed the threshhold value away from the mean and are less than the mean correspond to dark foreground features. To extract bright foreground features, we select the maximum pel value from the two images using a MAX operator. Likewise, to extract dark foreground features, we apply a MIN operator to those foreground pels. To combine background features of two images, we average the corresponding background pels. There are instances when certain pel areas of one image contains bright foreground while corresponding pel areas in the other image contains dark foreground. In such cases, we average the pels. Figure 6 illustrates a decision matrix that shows what operator to use in the different scenarios of mixing two images. An example of mixing two images using this algorithm is illustrated in Figure 7. It is observed that the characteristic features of the two images have been combined into the mixed image. For mixing three images, we have a similar three-dimensional decision cube. In our application of multi-directional interpolation, this mixing technique is used to mix together blocks outputted from the directional interpolators.

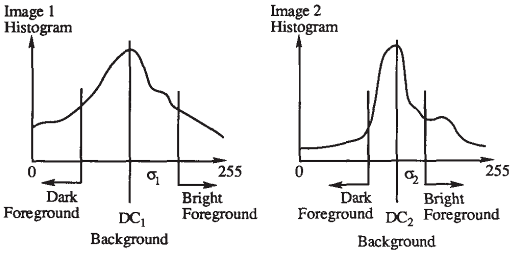


Figure 5. Histogram and statistics of two images to be mixed.

		Image 1		
		Bright Foreground	Background	Dark Foreground
Image 2	Bright Foreground	MAX	MAX	AVG
	Background	MAX	AVG	MIN
	Dark Foreground	AVG	MIN	MIN

Figure 6. Decision matrix for mixing two images.

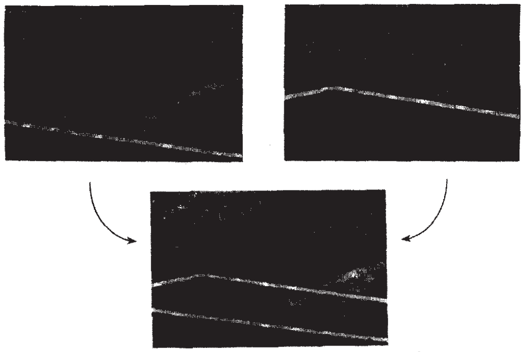


Figure 7. Example of mixing together two images.

Figure 8. "Flowergarden" image. (a) Damaged with isolated lost blocks (10.26 dB). (b) Restored using multi-directional interpolation (23.92 dB).

4. Experimental Results and Discussion

In the computer simulations, a number of still images and image sequences have been used. Results show very

Figure 9. "Flowergarden" image. (a) Damaged with consecutive lost blocks (13.31 dB). (b) Restored using multi-directional interpolation (26.12 dB).

good restoration if the missing block is surrounded by a neighborhood of pels with distinctive edges. When strong edges are present, the classifier works well in determining which directions to use in the interpolation. Figure 8(a) shows the 720 x 480 pel "Flower Garden" image with 25% of its 16 x 16 pel blocks lost. In this case, the lost blocks are assumed to be isolated so that there exists eight neighboring good blocks for each missing block. Figure 8(b) shows the restored image after applying the multi-directional interpolation algorithm to each of the missing blocks. It is observed that the restoration has a very good subjective viewing quality, with virtually no blocky artifacts. The restored blocks seem to be visually consistent with its neighboring pels. The lamp post's single strong edge has been restored; many of the lost blocks which damaged the windmill and tree branches have been restored with multiple interpolated edges. The texture of the large tree trunk and flowers on the ground has also been restored successfully. The peak signal-to-noise ratios (PSNRs) of Figures 8(a) and 8(b) are 10.26 dB and 23.92 dB respectively.

In most video block-coding schemes, however, lost blocks are rarely isolated from one another because video image blocks are often transmitted in raster-scan order. The MPEG international video coding standard [1] uses a raster-scan ordering of blocks to achieve a gain in compression by differentially coding dc values and motion vectors. When an uncorrectable error occurs in an MPEG bitstream, several consecutive adjacent blocks are usually lost until synchronization can be re-established. Figure 9(a) illustrates a drastic example of such errors occuring in one frame of the "Flower Garden" video sequence. The only available good neighboring pels come from above and below each of the missing blocks. Figure 9(b) demonstrates the capability of the concealment algorithm under such conditions. The restored image still looks resonably good, especially in areas with edges oriented close to vertical. Because of the nonexistence of neighboring good pels on either side of the missing block, the restoration cannot be expected to perform as well in restoring edges with directions close to horizontal. This effect can be seen at the base of the lamp post in Figure 9(b), whereby the horizon could not be restored. Overall, the subjective quality of the restored image is good, even when viewed as a still image. There is very little blurring and blocky artifacts present. The PSNRs of Figures 9(a) and 9(b) are 13.31 dB and 26.12 dB respectively.

In the video sequence simulations, quantitative results are obtained by calculating PSNR as a function of frame number. The video sequence source tested is "Cheerleaders" which contains moderately irregular motion throughout; it consists of 704 x 480 pixel image frames and is encoded with an MPEG encoder at 7 Mbps. MPEG coding parameters N=15 and M=3 were used, which effectively limits the extent of temporal error propagation to be less than 15 frames. The resulting bitstream is then packetized into 47-byte ATM-style data packets and errors were introduced with a cell loss probability of 10^{-3} and mean burst length of 2. In the MPEG context, a *slice* consists of an integral number of *macroblocks* grouped together in raster-scan row order. Each MPEG macroblock consists of

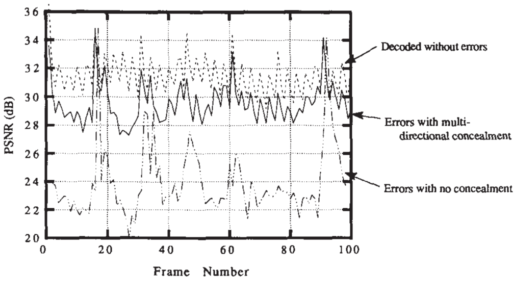


Figure 10. Video experiment on "Cheerleaders" sequence.

the combination of 16 x 16 pixel luminance blocks and corresponding chrominance blocks. Upon occurrence of a lost packet, resynchronization into the bitstream begins at the start of the next slice. A slice size consisting of 11 macroblocks was chosen to localize the errors. We compare the performance for three different cases: 1) decoded sequence without errors 2) decoded sequence with errors and no concealment 3) decoded sequence with errors and multi-directional interpolation concealment algorithm. The no-concealment case is obtained by filling in lost pixel areas with zeroes. Figure 10 shows the results of cases (1), (2), and (3) plotted on the same set of axes. Figure 11 shows the average PSNR for all three cases. The PSNRs are computed relative to the original uncoded source images. It is observed that the multi-directional interpolation concealment algorithm improves the non-concealed images by approximately 5.7 dB and is within 2 dB of perfectly decoded images.

Simulation Case	PSNR (dB)
Decoded Without Errors	31.20
Multi-directional Concealment	29.45
No Concealment	23.70

Figure 11. Average PSNR for the "Cheerleaders" sequence.

5. Conclusions

In this paper, an algorithm using multi-directional interpolation has been presented for error concealment. In the algorithm, the edge classifier analyzes the values of pels in the blocks surrounding the missing block and determines which edge directions cut through the missing block. Interpolations are performed on a local pel neighborhood along the directions specified by the edge classifier. Then a mixing operation is used to restore the missing block by extracting the features obtained from the different directional interpolations and combining them together. The method of multi-directional interpolation and image mixing has demonstrated very good results when there exists a sufficiently large neighborhood of correlated pels. This method of spatial interpolation can be combined with temporal interpolation to provide for a powerful error concealment technique for compressed video signal transmission.

References

[1] ISO-IEC/JTC1/SC29/WG11, "Coding of Moving Pictures and Associated Audio," MPEG 1992.

[2] Y. Wang and Q. Zhu, "Signal Loss Recovery in DCT-Based Image and Video Codecs," SPIE Proceedings on Visual Communication and Image Processing '91, Nov. 11-13, 1991, Boston, MA, pp.667-678.

[3] H. Sun, K. Challapali, and J. Zdepski, "Error Concealment in Simulcast ADTV Decoder," Proceedings of ICCE, June 1992.

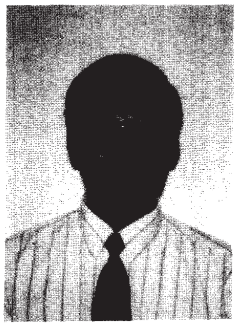
[4] H. Sun and W. Kwok, "Concealment of Damaged Block Transform Coded Images Using Projections onto Convex Sets," Submitted to IEEE Trans. Image Proc., April 1993.

[5] R. Siracusa, K. Joseph, J. Zdepski, and D. Raychaudhuri, "Flexible and Robust Packet Transport for Digital HDTV," IEEE JSAC, Vol 11 No. 1, Jan. 1993.

[6] E.R. Davies, *Machine Vision*, Academic Press Inc., San Diego, 1990.

[7] B. Ramamurthi and A. Gersho, "Nonlinear Space-Variant Postprocessing of Block Coded Images" IEEE Trans. ASSP, Vol. ASSP-34 No. 5, pp. 1258-1267, Oct. 1986.

About the Authors


[redacted]

Wilson Kwok was born in Brooklyn, NY in 1970. He received the B.S. and M.S. degrees in Electrical Engineering, summa cum laude, from The Cooper Union, New York, NY, in 1991 and 1992, respectively. He was the recipient of the IEEE Region I student paper award in 1991 for his work on synthetic aperture radar imaging systems. He joined the David Sarnoff Research Center in 1992, and currently works on MPEG / HDTV development. His research interests include radar imaging systems, image/video processing, compression, restoration and recognition.


[redacted]

Huifang Sun received the B.S. degree in electrical engineering from Harbin Engineering Institute, Harbin, China in 1967, and the Ph.D. degree in electrical engineering from University of Ottawa, Ottawa, Canada in 1986. Since 1970 he was an Engineer at the Shanghai Aroenautical Radio and Electronic Research Institute, Shanghai, China. In 1981 he was a visiting scholar at Laval University, Quebec, Canada. From 1982 to 1986, he was with the digital image processing laboratory in electrical engineering at University of Ottawa where he served as a Teaching Assistant, Reaserch Assistant and Post-doctoral Fellow. In 1986 he jointed Fairleigh Dickinson University, Teaneck, New Jersey, as an assistant professor and later as an associate professor in electrical engineering. Since 1990 he has been with the David Sarnoff Research Center (formerly RCA Laboratories), Princeton, New Jersey, as a member of technocal staff. His current interests include digital image processing, video data compression and error concealment, visual communication and HDTV design.