AN ADAPTIVE MARKOV RANDOM FIELD BASED ERROR CONCEALMENT METHOD FOR VIDEO COMMUNICATION IN AN ERROR PRONE ENVIRONMENT

Shahram Shirani, Faouzi Kossentini, and Rabab Ward

Department of Electrical and Computer Engineering,
University of British Columbia,
Vancouver BC V6T 1Z4, Canada

ABSTRACT

Loss of coded data during its transmission can affect a decoded video sequence to a large extent, making concealment of errors caused by data loss a serious issue. Previous work in spatial error concealment exploiting MRF models used a single pixel wide region around the erroneous area to achieve a reconstruction based on an optimality measure. This practically restricts the amount of available information that is used in a concealment procedure to a small region around the missing area. Incorporating more pixels usually means a higher order model and this is expensive as the complexity grows exponentially with the order of the MRF model. Using previously proposed approaches, the damaged area is reconstructed fairly well in very low frequency portions of the image. However, the reconstruction process yields blurry results with a significant loss of details in high frequency, or edge portions of the image. In our proposed approach, a MRF is used as the image a priori model. More available information is incorporated in the reconstruction procedure not by increasing the order of the model but instead by adaptively adjusting the model parameters. Adaptation is done based on the image characteristics determined in a large region around the damaged area. Thus, the reconstruction procedure can make use of information embedded in not only immediate neighborhood pixels but also in a wider neighborhood without a dramatic increase in computational complexity. The proposed method outperforms the previous methods in the reconstruction of missing edges.

1. INTRODUCTION

The fast growth of digital transmission services has created a great interest in digital transmission of image and video signals. Since digital image and video signals require very high bit rates, the compression of such signals becomes necessary. Three standards have emerged to facilitate the growth of new image communication applications. These are: 1) the Joint Photographic Experts Group (JPEG) standard for still image compression [1, 2], 2) the International Telecommunication Union (ITU) recommendation H.261 for video telephony and conferencing [1] and its subsequent revisions, e.g., H.263 and H.263+, and 3) the Moving Pictures Expert Group (MPEG) for full motion video compression and coding in digital storage media and digital communication applications [1]. The common features of these compression standards is that they are all block based and use the discrete cosine transform (DCT).

Communication channels are not error free and consequently, the encoded bit streams are vulnerable to transmission impairments, which may cause loss of blocks of data or total loss of synchronization. The impact of bit stream corruption or loss in the picture quality while usually substantial, still depends on the actual transmission method and the compression algorithm.

Several methods to combat the channel errors have been proposed. Automatic Retransmission Request (ARQ) or interleaving techniques are often ineffective, because ARQ may aggravate channel congestion and cause the system to drop more data, and interleaving may require considerable delay. Another method is to employ forward error control coding techniques. There are several problems associated with these techniques. First, they usually require too many additional precious bits for error detection and/or correction. Second, they may introduce long delays that are not acceptable in some applications. An alternative method is to perform error concealment, which intends to conceal the bit error effects at the receiver by exploiting redundancies in the video signal and limitations of the human visual system, without requiring additional information at the coder [3, 4]. Error concealment of images and video aims at removing the visually annoying artifacts that degrade significantly the overall picture quality.

Error concealment techniques mainly make use of temporal and/or spatial correlation in the video signal and reconstruct the missing region of video frame from adjacent regions or frames. A simple and yet quite effective temporal reconstruction method is to replace the corrupted/missing region with its corresponding part in the previous frame. Although this method generally works well in still parts of the picture, such as the background, it cannot produce satisfactory results when the video sequence exhibits fast moving objects, lighting changes, or sudden scene changes[5]. Spatial reconstruction techniques include averaging or linear interpolation [6], constrained linear interpolation [6], multi-directional edge-based interpolation [7, 8, 4], and Bayesian interpolation [9].

Intuitively, the most effective reconstruction method is the one that uses the image a priori model together with the available data for estimating the missing data. Recently, Markov Random Fields (MRF) have been extensively used to model images. The attractive features of an MRF model are the computational tractability and the ability of the model to capture non-Gaussian aspects of the image such as edges.

Image distribution models (e.g., MRF) have been used for error concealment. In fact, the Bayesian approach provides a frame work for incorporating the a priori information through the choice of the distribution of the image. Maximum a priori (MAP) estimation, a paradigm used very often in image processing, yields the most likely image given the observed data. A critical component in MAP estimation is the prior distribution of the image model.

Previous work in spatial error concealment exploiting MRF

This work was supported by MDSI, BC ASI and NSERC.

models used a single pixel wide region around the erroneous area to achieve a reconstruction based on an optimality measure. In other words, the amount of available information that is used in estimating damaged data, has been restricted. Incorporating more pixels usually means a higher order MRF model and this is expensive as the complexity grows exponentially with the order of the MRF model. Using previously proposed approaches, the damaged area is reconstructed fairly well in very low frequency portions of the image. However, the reconstruction process yields blurry blocks with a significant loss of details in high frequency or edge portions of the image.

In our approach, a MRF is used as the image a priori model. We incorporate more available information in our suggested reconstruction method not by increasing the order of the model but instead by adaptively adjusting the model parameters. Adaptation is done based on the image characteristics determined using a large region around the damaged area. Thus, the reconstruction procedure can make use of information embedded in not only immediate neighborhood pixels but also in a wider neighborhood without a dramatic increase in computational complexity. The proposed method achieves better performance in reconstructing the edges. Although the method is general and can be applied to any of block-based compression method (for images or image sequences) we use H.263 video coding method.

It is assumed throughout that the decoder knows the locations of the missing macroblocks (MB). This can be achieved, for example, by transmitting the MBs of the image sequence in a predetermined order and assigning sequence numbers to packets in packet based transmission.

The structure of this paper is as follows. In Section 2, MRF and MAP estimation of missing data is discussed. Section 3 details the proposed method. Sections 4 and 5 present the experimental results and the conclusions, respectively.

2. MRF AND MAP ESTIMATION OF THE MISSING DATA

Over the last few years, MRFs have been extensively used for image modeling [10]. The attractive features of an MRF model are computational tractability and the ability of the model to capture non-Gaussian aspects of the image such as edges. To enable the model to accurately characterize the image data, usually adjustable parameters are considered in the model. A MRF with a Gibbs distribution is

$$Pr(x) = \frac{1}{Z} exp\{-\sum_{c \in C} V_c(x)\}, \tag{1}$$

where Z is a normalizing constant, $V_c(.)$ is called potential function and is a function of a local group of pixels c called clique, and C denotes the set of all cliques throughout the image. Depending on the choice of the potential function and the clique, different models are obtained. Each potential function characterizes the relationship between a group of pixels by assigning a larger cost to configurations of pixels which are less likely to occur.

Having selected the image model, estimation of image missing data using the MAP estimation technique eventuates into a functional minimization problem [9, 11]. The choice of the potential function is crucial to the quality of the estimated image. The potential function should be convex to have an easily-obtainable global minimum. Commonly, the potential functions are selected to be in the form of

$$\sum_{c \in C} V_c(x) = \sum_{c \in C} \rho(\mathbf{d_c^t x}), \tag{2}$$

where $\mathbf{d_c}$ is a coefficient vector for a clique c. These coefficients are selected based on the a priori assumptions about the image. Usually they are selected so that $\mathbf{d_c}$ provides an approximation of first or second derivative of the image at each pixel. We will consider only potential functions which act on pairs of pixels. For the special case of $\rho(x) = x^2$, the model is called a Gauss-Markov random field (GMRF). This image model may result in blurred estimate of the image particularly along edges. To reduce the smoothing effect of the GMRF other forms of cost functions have been introduced. One of the proposed cost functions is the Huber function

$$\rho(x) = \begin{cases} x^2 & : \ |x| < \gamma \\ \gamma^2 + 2\gamma(|x| - \gamma) & : \ |x| > \gamma \end{cases} \tag{3}$$

where γ is the threshold. The image model exploiting this cost function is called a Huber Markov Random Field (HMRF).

In this work, we will consider an eight pixel clique around each pixel as shown in Figure 1. There are eight directions corresponding to the line segment connecting the pixel and one of the pixels in its clique. The potential function can be written as

$$\sum_{c \in C} V_c(x) = \sum_i \sum_j \sum_{m=1}^{8} \rho(w_{i,j}^m D_m(x_{i,j})), \tag{4}$$

where $D_m(x_{i,j})$ is the difference between the value of the pixel in position (i,j) and the pixels in its clique corresponding to m-th direction and $w_{i,j}^m$ is the weight assigned to this difference. The reason for selecting an eight pixel clique in the way shown will become clear in the following section.

3. PROPOSED METHOD

Basically, the performance of the MAP estimator based on a HMRF can be described in this way: when the values of the neighboring pixels are close to each other, the missing pixel is set to the average of those pixels. When the pixel values are not similar, a voting procedure is performed and the estimated value is selected that is close to the value of the majority of the neighboring pixels (a median like performance). This behavior prevents the appearance of pixel values different from their neighbors, which in turn limits the performance of the estimator in reconstructing edges. A very simple situation is depicted in Figure 2. The value of the center pixel, $x_{i,j}$ is missing. Assuming the pixel values are p and q as shown in the figure and $p \ll q$, there will be a vertical edge in the image. Using the GMRF model, the value of the missing pixel is

$$\hat{x}_{i,j} = \sum_{(i,j) \in c \cup c'} x_{i,j} / (n_c + n_{c'}), \tag{5}$$

where c is the clique and c' is its complement shown in Figure 1, n_c and $n_{c'}$ are the number of pixels in the clique and its complement respectively (each 8 for the shown clique) [11]. For the shown values, the estimation will be

$$\hat{x}_{i,j} = \frac{2p + 14q}{16}, \tag{6}$$

Exploiting the HMRF model we will get

$$\hat{x}_{i,j} = (\sum_{(i,j)\in I} x_{i,j} - 2\gamma)/n_I, \quad (7)$$

where $I = c \cup c' - \{(i-1,j),(i+1,j)\}$ and n_I is the number of pixels in I. For the specific example, we will have

$$\hat{x}_{i,j} = \frac{14q - 2\gamma}{14} \quad (8)$$

Obviously, the estimated value is close to the majority of the neighboring pixels. Thus, none of the above mentioned models is able to detect the presence of the vertical edge and reconstruct the missing pixel value based on that edge. Usually, relying only on the local image characteristics (e.g. using a window) in the reconstruction procedure causes some of the image attributes to be ignored or misinterpreted.

In this work, instead of using the HMRF which seems to be ineffective, we exploit the GMRF model with an eight pixel neighborhood as the clique. The weight corresponding to the differences between a pixel and each of the pixels in its clique is selected adaptively, based on the likelihood of an edge in the direction of that pair of pixels. The rationale behind this selection is to weigh more the difference between the pixels in that direction. This will cause the values of the pixels in that direction to get closer to each other. The likelihood of edges in each of the eight directions is calculated using a large area around the missing MB. In this way, the available information is exploited on a larger area without increasing the order of GMRF model which consequently increases computational complexity. As the weights are selected according to the edges in the corresponding direction, when several edge directions exist, the reconstruction procedure combines them.

The first step in the proposed method is to determine those edges in the MBs surrounding the missing MBs that pass through the missing MBs. Edges in the MBs to the left, right, up and down of the missing MB are determined using the gradient measure in the spatial domain [7, 8, 4]. The edge for pixel x(i,j) in the surrounding MBs is computed by

$$g_x = x_{i+1,j} - x_{i-1,j-1} + x_{i+1,j} \quad (9)$$
$$-x_{i-1,j-1} + x_{i+1,j} - x_{i-1,j-1},$$
$$g_y = x_{i+1,j} - x_{i-1,j-1} + x_{i+1,j} \quad (10)$$
$$-x_{i-1,j-1} + x_{i+1,j} - x_{i-1,j-1}$$

which is the Sobel mask. The magnitude and angular direction of the gradient at (i,j) are:

$$G = \sqrt{g_x^2 + g_y^2} \quad (11)$$
$$\theta = \arctan(\frac{g_y}{g_x}) \quad (12)$$

The angular value of the gradient is rounded to one of the eight directions equally spaced in the zero to $180°$. There is a counter corresponding to each of the eight directions. If a line drawn through the pixel (i,j) passes thought the missing block, the corresponding counter is incremented by the amount of the gradient. The weights in the potential function corresponding to each pair of pixels is selected proportional to the counter in the direction corresponding to them. That is

$$w^m_{i,j} = \alpha c_m \quad (13)$$

where c_m is the counter in the m-th direction and α is a constant. It can be shown along the proof given in [11] that under these conditions

$$\hat{x}_{i,j} = \sum_{(i,j)\in c\cup c'} w^m_{i,j} x_{i,j} / (\sum_{(i,j)\in c\cup c'} w^m_{i,j}), \quad (14)$$

Therefore an iterative procedure can be exploited to estimate the missing pixel values. Finally, the whole reconstruction procedure can be described as follows:

1. Determine the edges in the neighbor MBs and assign them to eight equally spaced directions. Count the number of edges in each direction,

2. Assign a value proportional to each edge counter to the corresponding weight in the GMRF model,

3. Use (14) to find an estimation of each missing pixel based on its adjacent pixels and the weight obtained in the previous steps, and

4. Iteratively estimate the missing pixels using (14) until the procedure converges.

In the case where adjacent MBs are lost, the reconstruction algorithm is applied recursively starting with the MBs with maximum number of correctly decoded neighbors.

4. EXPERIMENTAL RESULTS

The proposed error concealment method has been tested using a H.263 video coder. Figure 3 shows a frame of the image sequence FOREMAN coded and decoded using H.263. Figure 4 shows the same frame missing a number of MBs. Figure 5 shows the image reconstructed using the non-adaptive GMRF model. Obviously, this method performs poorly in reconstruction of edges. Figure 6 shows the image of Figure 4 after applying the proposed error concealment algorithm. Clearly, the proposed method is performing better in retrieving the edges. The PSNR for reconstructed frames using non-adaptive GMRF and adaptive GMRF are 28.1 and 31.8 respectively. The results shown here, as well as results obtained for other test images, demonstrate that the proposed algorithm perform well.

5. CONCLUSIONS

In this paper we presented a new approach for reconstruction of missing coded data. In the suggested method, a MRF is used as the image a priori model and the model parameters are adaptively and locally adjusted based on the image characteristics determined using a large region around the damaged area. In this way, the reconstruction procedure exploits the information embedded in a large neighborhood around the area with missing data without a substantial increase in computational complexity. The missing data is estimated using a MAP estimator and the adaptive MRF model. The proposed method outperforms previous methods in reconstructing the edges and the quality of the reconstructed images is also relatively good.

6. REFERENCES

[1] K. R. Rao and J. J. Hwang, *Techniques and Standards for Image, Video and Audio Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

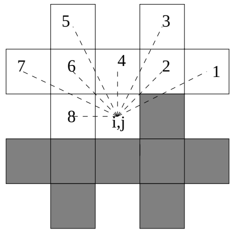

Figure 1: A pixel, its clique c and the eight directions. c' is the dark area.

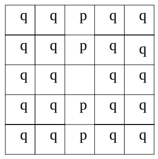

Figure 2: A missing pixel in a vertical edge.

[2] ISO/IEC, *Digital compression and coding of continuous tone still images: requirements and guidlines*. International Standard Organization, 1994.

[3] Y. Wang and Q. F. Zhu, “Error control and concealment for video communication: A review,” *Proceedings of the IEEE*, pp. 974–997, May 1998.

[4] Q. F. Zhu, Y. Wang, and L. Shaw, “Image reconstruction for hybrid video coding systems,” in *Proc. IEEE Data Compression Conf.*, (Snowbird, UT), pp. 229–238, Mar. 1992.

[5] M. Ghanbari and V. Seferidis, “Cell loss concealment in ATM video codecs,” *IEEE Transactions on Circuit and system in video technology*, vol. 3, pp. 238–247, June 1993.

[6] P. Salama, N. B. Shroff, E. J. Coyle, and E. J. Delp, “Error concealment techniques for encoded video streams,” in *Proceedings of International Conference on Image Processing*, pp. 9–12, 1995.

[7] H. Sun and W. Kwok, “Concealment of damaged block transform coded images using projection onto convex sets,” *IEEE Transactions image processing*, vol. 4, pp. 470–477, April 1995.

[8] W. Kwok and H. Sun, “Multi-directional interpolation for spatial error concealment,” *IEEE Transactions on consumer electronics*, vol. 39, pp. 455–460, August 1993.

[9] P. Salama, N. B. Shroff, and E. J. Delp, “A baysian approach to error concealment in encoded video streams,” in *Proceedings of International Conference on Image Processing*, pp. 49–52, 1996.

[10] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images,” *IEEE Trans. Pattern Anal. and Mach. Int.*, vol. 11, no. 6, pp. 689–691, 1984.

[11] P. Salama, N. B. Shroff, and E. J. Delp, “A fast suboptimal approach to error concealment in encoded video streams,” in *Proceedings of International Conference on Image Processing*, 1997.

Figure 3: Original frame of the image sequence Foreman.

Figure 4: Frame missing MBs.

Figure 5: Frame after reconstruction using a GMRF model.

Figure 6: Frame after reconstruction using our adaptive GMRF model.