

SPATIALLY ADAPTIVE FILTERING FOR VIDEO ENCODING

5

Field of the Invention

The present invention relates generally to video coding, and more particularly to utilizing a spatially adaptive preprocessing filter for use in video coding.

10

Background of the Invention

Current standards, such as H.261, MPEG1, and MPEG2 provide compression of a digital video sequence by utilizing a block motion-compensated Discrete Cosine Transform, DCT, approach. This video encoding technique removes the redundancy present in a video sequence by utilizing a two-step process. In the first step, a block-matching, BM, motion estimation and compensation algorithm estimates the motion that occurs between two temporally adjacent frames. The frames are then compensated for the estimated motion and compared to form a difference image. By taking the difference between the two temporally adjacent frames, all existing temporally redundancy is removed. The information that remains is new information that was not compensated for in the motion estimation and compensation algorithm. This difference image is referred to as the displaced frame difference, DFD.

2

In the second step, the DFD image is transformed into a frequency domain using the DCT. The DCT has the property of compacting the energy of the DFD image into a few low frequency components. Further compression of the video sequence is obtained by limiting the amount of high frequency information encoded.

5

If noise is present in a video sequence, then the DFD image will also contain the difference of this noise. Since the noise is uncorrelated in the temporal direction, the energy of the noise present in the DFD image is actually double the noise energy present in any single frame. After applying the DCT, the noise present in the DFD results in an increase in the energy present in the high frequency components. The net effect is the compression technique is forced to use a large number of bits, as compared to the case when the noise is not present, to maintain a particular distortion in the decoded video. In other words, noise present in a video sequence prior to encoding, results in a large decrease in the compression efficiency of any compression technique.

10
15

Current approaches for removing noise from a video sequence are inadequate for two reasons. The first reason is that these approaches attempt to remove the noise from the intensity frames of the video sequence. Generally, this type of approach introduces artifacts into the video sequence that are then transferred to the DFD signal and interpreted by the video encoder as new information that must be encoded. In other words, these approaches can also reduce the compression efficiency of the video encoder. The second reason that current approaches are inadequate is that they tend to operate only in the temporal direction. That is, they

20

3

filter along a single direction between frames. If averaging is used, then the parts of the video sequence that are in motion will be blurred. By blurring the moving objects the video sequence in general will be easier to encode. However, it will look substantially degraded to the viewer. Therefore, the current approaches to removing noise are unable to reduce the amount of information that must be encoded by the video encoder without substantially degrading the decoded video sequence.

Brief Description of the Drawings

10

FIG. 1 is a block diagram of a preferred embodiment of an apparatus for spatially adaptive filtering for video encoding in accordance with the present invention.

15

FIG. 2 is a flow diagram of a preferred embodiment of steps of a method for spatially adaptive filtering for video encoding in accordance with the present invention.

Detailed Description of Preferred Embodiments

20

The method and apparatus described below filter a video sequence prior the encoding process of a video sequence. Specifically, a method and apparatus are described wherein a spatially adaptive filter is used to remove noise directly from the Displaced Frame Difference, DFD, signal. This novel approach removes noise and

miscellaneous high frequency components from the DFD signal without the introduction of the filtering artifacts characteristic of current techniques. By reducing the miscellaneous high frequency components the present invention is capable of reducing the amount of information that must be encoded by the video encoder
5 without substantially degrading the decoded video sequence.

In order to describe the method and apparatus certain assumptions are made concerning the input video sequence. Specifically, the video source is assumed to be in a digital format where the number of pixels per row, the number of rows per frame,
10 and the number of frames per second are known prior to processing. Each pixel represents both the luminance and chrominance components using 8 bit integer numbers which span from 0 to 255. As mentioned above, these assumptions are only made to help facilitate the description of the method and apparatus and should not be viewed as restrictions to applications where these assumptions do not hold.

15

FIG. 1, numeral 100 is a block diagram of a preferred embodiment of an apparatus for spatially adaptive filtering for video encoding in accordance with the present invention. The apparatus comprises a noise variance determiner (102), local variance determiner (104), noise visibility function determiner (106), Gaussian kernel
20 determiner (108), and a convolver (110). Utilizing a DFD (120) signal, the noise variance determiner (102) calculates an estimate of the noise variance $\sigma_n^2(r)$ (126). Also utilizing the DFD (120) signal the local variance determiner determines an estimate of the local signal variance $\sigma_f^2(r)$ (128). Based on the local signal variance $\sigma_f^2(r)$ (128) and noise variance $\sigma_n^2(r)$ (126) estimates and a predetermined image

type parameter Θ (130), the noise visibility function determiner determines a noise visibility parameter $\nu(\Theta)$ (132) which controls the amount of filtering provide by the spatially adaptive preprocessing filter at pixel location $(r) \equiv (i, j)$. Specifically, the noise visibility parameter $\nu(\Theta)$ (132) is used by the Gaussian kernel determiner (110) to determine a Gaussian filtering kernel that will provide the proper amount of local filtering based on the local SNR.

The apparatus may be embodied in computer software, a Digital Signal Processor, a gate array, or an application specific integrated circuit.

10

Utilizing the DFD signal, the noise variance determiner (102) determines an estimate of the noise variance σ_n^2 (126). This is accomplished by first dividing the DFD signal into 16×16 blocks and calculating the energy contained in these blocks. It should be noted that if block matching, BM, is used in generating the DFD signal, then these energy calculations already exist. The energy is calculated using the following expression

15

$$\sigma_B^2(B_m) = \frac{1}{MN} \sum_{i,j \in B_m} (DFD(i,j))^2, \quad (1)$$

20 where pixel location i, j is a member of block B_m , M and N are the number of pixels contained in a row and column of block B_m , and $\sigma_B^2(B_m)$ is the energy contained in block B_m . Once the energy, $\sigma_B^2(B_m)$, is calculated for each block they are rank

6

ordered from highest to lowest energy. The average of the bottom 25% values in this ranking is used as the estimate of the noise variance, σ_n^2 (126).

The local signal variance is calculated utilizing the noise variance estimate σ_n^2 (126) and a local variance $\sigma_f^2(r)$. It is known that the local DFD variance $\sigma_f^2(r)$ is the total of the local signal variance $\sigma_s^2(r)$ (128) plus the noise variance σ_n^2 (126). Therefore, an estimate of the local signal variance is obtained by subtracting the noise variance estimate from the local DFD signal variance. This is demonstrated below

$$\sigma_s^2(r) = \sigma_f^2(r) - \sigma_n^2(r) \quad , \quad (2)$$

where

$$\sigma_f^2(r) = \frac{1}{MN} \sum_{\substack{-K \leq k \leq K \\ -L \leq l \leq L}} (DFD(i+k, j+l))^2 \quad , \quad (3)$$

K and L define the local neighborhood that the local signal variance is calculated over. Specifically, the local signal variance $\sigma_s^2(r)$ (128) is calculated within a block that extends K rows above and below pixel $(r) \equiv (i, j)$, as well as L columns to the left and right.

The visibility parameter $v(\Theta)$ (132) which controls the amount of filtering provide by the spatially adaptive preprocessing filter at pixel location $(r) \equiv (i, j)$, is calculated by the noise visibility function determiner. Based on the local signal variance $\sigma_s^2(r)$ (128) and noise variance $\sigma_n^2(r)$ (126) estimates and a predetermined

7

image type parameter Θ (130), the noise visibility function determiner determines the noise visibility parameter using the following equation

$$v(\Theta) = \frac{1}{1 + \frac{\sigma_f^2}{\sigma_n^2} \Theta}. \quad (4)$$

- 5 To reduce noise present in a DFD signal the predetermined image type parameter Θ (130) is typically set at a value of 0.01. For different types of image data the image type parameter Θ is adjusted to provide the visibility with a different response. For example, due to their sensitivity to over-smoothing as opposed to the presence of noise, intensity images would require the image type parameter Θ to
 10 increased when compared to the value used for a DFD signal.

- Based on the noise visibility parameter $v(\Theta)$ (132) the Gaussian kernel determiner (110) determines a Gaussian filtering kernel that will provide the proper amount of local filtering based on the local SNR. The response of the Gaussian
 15 kernel determiner (110) to the noise visibility parameter $v(\Theta)$ (132) is shown below

$$g(m,n) = \frac{1}{Z} e^{-\left[\frac{m^2+n^2}{v(\Theta)\sigma_f^2} \right]}, \quad (5)$$

- 20 where $g(m,n)$ is the Gaussian filtering, (m,n) is a location in the filters support, and σ_f^2 is the variance of the filter. The support of the filter is determined prior to implementation. Typically, to reduce the number of operations required to be

performed in the convolver (110), the support of the filter is 5 pixel rows by 5 pixel columns. The variance of the filter is also predetermined and along with the noise visibility parameter $v(\Theta)$ (132) is used to control the amount of smoothing applied to the DFD by the filter.

5

The noise visibility parameter $v(\Theta)$ (132) controls the amount of smoothing applied to the DFD by the filter, in the following way. If the noise variance $\sigma_n^2(r)$ (126) is small compared to the local signal variance $\sigma_f^2(r)$ (128), then the noise visibility parameter $v(\Theta)$ (132) is also be small. This results in a Gaussian filter $g(m,n)$ with a narrow spatial response or a wide pass-band in the frequency domain. In other words, the filter $g(m,n)$ would provide very little smoothing. This is the desired response, since the noise variance is small compared to the signal variance. For the case when the local signal variance is much smaller than the noise variance, $g(m,n)$ will have a wide spatial response which would result in a narrow passband in the frequency domain. Again, this is the desired response, since the $g(m,n)$ would provide smoothing to a very noisy DFD.

10

15

20

The convolver (110) actually carries out the filtering operation on the DFD signal. That is, the convolver (110) applies the Gaussian filter $g(m,n)$ (138) to the DFD signal (120) through a convolution operation. The convolution operation is characterized by the following equation.

$$DFD(i,j) = \sum_{\substack{m=-5 \\ n=0 \\ m=5}}^{m=5, n=5} g(m,n) DFD(i-m, j-n), \quad (6)$$

where $D\hat{F}D$ (140) is the preprocessed DFD signal.

FIG. 2, numeral 200 is a flow diagram of a preferred embodiment of a method for spatially adaptive filtering for use in video encoding. The first step is calculating
5 the noise variance based on the DFD signal (202). Next, a local signal variance is determined based on the calculated noise variance and the DFD signal (204). Utilizing the local signal variance, noise variance, and a predetermined image type parameter, a noise visibility parameter is then calculated (206). A local Gaussian filter is determined based on the noise visibility parameter and a predetermined
10 Gaussian filter variance parameter (208). Then, the local Gaussian filter is convolved with the DFD signal to produce the filtered DFD signal (210).



10

Claims

We claim:

- 5 1. A method of spatially adaptive filtering a displaced frame difference, a DFD, the DFD is a difference between a video frame and an estimated video frame, the method comprising:

10 determining, using a noise variance determiner, a noise variance of the DFD based on a predetermined signal-to-noise ratio;

determining, using a local variance determiner, a local variance of the DFD based on the noise variance and the predetermined signal-to-noise ratio;

15 determining, using a noise visibility determiner, a noise visibility function based on the noise variance, the local variance, and a predetermined image type;

20 determining, using a Gaussian kernel determiner, a Gaussian kernel for every pixel in the DFD based on the noise visibility function, a predetermined normalizing coefficient, and a predetermined base variance; and

convolving, using a convolver, the DFD with the Gaussian kernel to provide an adaptively filtered DFD.



11

2. The method of claim 1 wherein at least one of 2A-2B:

2A) the noise variance is determined by calculating an energy contained in each predefined block using the following equation:

$$\sigma_B^2(B_m) = \frac{1}{MN} \sum_{(i,j) \in B_m} (DFD(i,j))^2,$$

5

where M and N are the dimensions of the predefined block and i, j denotes a row and column of a pixel in the predefined block; and

2B) the local variance of the DFD is determined using the following equation:

10

$$\sigma_d^2(r) = \frac{1}{MN} \sum_{\substack{-K \leq k \leq K \\ -L \leq l \leq L}} (DFD(i+k, j+l))^2$$

where M and N are the dimensions of the predefined block and i, j denotes a row and column of a pixel in the predefined block.

15

3. The method of claim 1 wherein at least one of 3A-3C:

3A) the noise visibility function is determined using the following equation:

20

$$v(\Theta) = \frac{1}{1 + \frac{\sigma_f^2}{\sigma_n^2} \Theta}$$



where σ_f^2 is the local variance, σ_n^2 is the noise variance, and Θ is the predetermined image type;

3B) the Gaussian kernel is determined using the following equation:

$$5 \quad g(m,n) = \frac{1}{Z} e^{\left[\frac{-m^2 + n^2}{v(\Theta)\sigma_s^2} \right]}$$

where σ_s^2 is the predetermined base variance, Z is the predetermined normalizing coefficient, $v(\Theta)$ is the noise visibility function, and m,n denotes a row and column of a pixel in the DFD; and

10 3C) the DFD convolved with Gaussian kernel is determined using the following equation:

$$D\hat{F}D(i,j) = \sum_{\substack{m=-5 \\ n=-5 \\ m=0 \\ n=0}}^{m=5} g(m,n) DFD(i-m, j-n),$$

15 where $g(m,n)$ is the Gaussian kernel and m,n denotes a row and column of a pixel in the DFD.

4. The method of claim 1 wherein at least one of 4A-4C:

4A) the steps of the method are embodied in a tangible medium of/for a
20 Digital Signal Processor, DSP;

4B) the steps of the method are embodied in an Application Specific Integrated Circuit, ASIC; and



4C) the steps of the method are embodied in a tangible medium of/for gate array.

5. The method of claim 1 wherein the steps of the method are embodied in a
5 tangible medium of/for a computer, and where selected, wherein one of 5A-5B:

5A) the tangible medium is a computer diskette;

5B) the tangible medium is a memory unit of the computer.



6. An apparatus for spatially adaptive filtering a displaced frame difference, a DFD, the DFD is a difference between a video frame and an estimated video frame, the apparatus comprising:

5 a noise variance determiner, operably coupled to receive the DFD and a predetermined signal-to-noise ratio, for determining a noise variance of the DFD based on the predetermined signal-to-noise ratio;

a local variance determiner, operably coupled to the noise variance determiner
10 and operably coupled to receive the DFD, for determining a local variance of the DFD based on the noise variance and the predetermined signal-to-noise ratio;

a noise visibility determiner, operably coupled to the noise variance determiner and the local variance determiner and operably coupled to receive a
15 predetermined image type, for determining a noise visibility function based on the noise variance, the local variance, and the predetermined image type;

a Gaussian kernel determiner, operably coupled to the noise visibility determiner and operably coupled to receive a predetermined base variance, for
20 determining a Gaussian kernel for every pixel in the DFD based on the noise visibility function, a predetermined normalizing coefficient, and a predetermined base variance; and



a convolver, operably coupled to the Gaussian kernel determiner and operably coupled to receive the DFD, for convolving the DFD with the Gaussian kernel to provide an adaptively filtered DFD.



16

7. The apparatus of claim 6 wherein at least one of 7A-7B:

7A) the noise variance is determined by calculating an energy contained in each predefined block using the following equation:

$$\sigma_B^2(B_m) = \frac{1}{MN} \sum_{(i,j) \in B_m} (DFD(i,j))^2,$$

5

where M and N are the dimensions of the predefined block and i, j denotes a row and column of a pixel in the predefined block; and

7B) the local variance of the DFD is determined using the following equation:

10

$$\sigma_d^2(r) = \frac{1}{MN} \sum_{\substack{-K \leq k \leq K \\ -L \leq l \leq L}} (DFD(i+k, j+l))^2$$

where M and N are the dimensions of the predefined block and i, j denotes a row and column of a pixel in the predefined block.

15 8. The apparatus of claim 6 wherein at least one of 8A-8C:

8A) the noise visibility function is determined using the following equation:

$$v(\Theta) = \frac{1}{1 + \frac{\sigma_f^2}{\sigma_n^2} \Theta}$$

20

where σ_f^2 is the local variance, σ_n^2 is the noise variance, and Θ is the predetermined image type;



17

8B) the Gaussian kernel is determined using the following equation:

$$g(m,n) = \frac{1}{Z} e^{-\left[\frac{m^2+n^2}{v(\Theta)\sigma_g^2} \right]}$$

5 where σ_g^2 is the predetermined base variance, Z is the predetermined normalizing coefficient, $v(\Theta)$ is the noise visibility function, and m,n denotes a row and column of a pixel in the DFD; and

8C) the DFD signal convolved with Gaussian kernel is determined using the following equation:

10

$$\hat{DFD}(i,j) = \sum_{\substack{m=-5 \\ n=0}}^{m=5} g(m,n) DFD(i-m, j-n),$$

where $g(m,n)$ is the Gaussian kernel and m,n denotes a row and column of a pixel in the DFD.

15 9. The apparatus of claim 6 wherein at least one of 9A-9C:

9A) the apparatus is in a Digital Signal Processor, DSP;

9B) the apparatus is in an Application Specific Integrated Circuit, ASIC;

and

9C) the apparatus is in a gate array.

20

10. The apparatus of claim 6 wherein the apparatus is embodied in a tangible medium of/for a computer, and where selected, one of 10A-10B:



- 10A) wherein the tangible medium is a computer diskette; and
- 10B) wherein the tangible medium is a memory unit of the computer.