PROJECT TOPIC:
A Study on structural similarity based interframe video coding

ABSTRACT:

The goal of the research is to study rate-distortion behavior of the interframe codec designed with novel motion estimation based on structural similarity distortion and the codec with conventional motion estimation based on pixel error distortion (absolute differences). The study from previous literature shows that structural similarity metric provides better image assessment than pixel error based metric (mean square error and peak signal-to-noise ratio). The codec with fixed motion block sizes (32×32), (16×16), (8×8), and (4×4), will be investigated. Simulation results such as rate-distortions will be measured in terms of both peak signal-to-noise ratio versus bit rate and structural similarity index versus bit rate.

MOTIVATIONS AND PROBLEM STATEMENTS:
In many video coding standards, including H.264/AVC [9, 10, 11], distortion calculation in motion estimation part is based on pixel error metric (for example, absolute differences, MSE). Pixel error metric is not designed to match with human visual perception. The disadvantages of error metric (MSE) can be found in [2, 3, 4, 6]. Structural similarity (SSIM) index is a new objective image quality metric, introduced by [3] for image assessment. In the research, we apply SSIM index for distortion calculation to get the motion vector during motion search process. The codec proposed in [1] will be used in the simulation. The reason is that it is less complicated to handle several components inside the codec. Rate-distortion results of different fixed-motion partition sizes will be observed. Perceived reconstructed video frame will be observed both at low bit rates and at high bit rates by varying quantization parameter. Recently, two papers [7, 8] introduced SSIM in motion estimation of H.264 codec.
METHOD:
The research uses the codec based on [1]. The codec is modified with single size motion estimation to observe the effect on each motion block size: (32×32), (16×16), (8×8), and (4×4). SSIM distortion is implemented in motion estimation part. Matlab implementation of SSIM can be obtained from [5]. Huffman codebooks for P-frame are trained with 5 test sequences. Simulations on 4 test video sequences with group of picture as IPP⋯ are performed. Rate-distortion result in terms of SSIM (video quality, Q, [4]) and bit rate is compared with conventional PSNR and bit rate. Low bit rate (low visual quality) and high bit rate (high visual quality) encoding regions will be observed and investigated.

BACKGROUND ON Definition of structural similarity index and the method of video quality assessment using SSIM:
This section is based on [3, 4]. Structural similarity (SSIM) index [3] is a new objective method to measure image quality between a distorted image and a reference image. The SSIM measures the degradation of structural information based on the assumption that human visual system characteristics are adopted for extracting structural information from an image scene. The authors claim that SSIM has better consistency with perceived image quality than pixel error model (absolute differences and MSE) based on different performance evaluations.

The author [3] provides the following observations on nature of natural images and human visual characteristics to implement SSIM locally.
1. Natural images are spatially non-stationary in general.
2. Image distortions can be space-variant. The distortions may or may not depend on local image statistics.
3. Only a local area in the image can be perceived by human at one time instance, regarding to HVS.
4. Localized quality measurement of the image provides a spatially varying quality map (called SSIM map) of the image which provides more quality degradation information of the image.
SSIM is defined as the product of three local quantities: luminance comparison (function of mean), contrast comparison (function of variance), and structure comparison (function of correlation coefficient and variance)

\[
\text{SSIM}(\bar{x}, \bar{y}) = \left[ l(\bar{x}, \bar{y}) \right]^\alpha \cdot \left[ c(\bar{x}, \bar{y}) \right]^\beta \cdot \left[ s(\bar{x}, \bar{y}) \right]^{\gamma}
\]

where,

\[
l(\bar{x}, \bar{y}) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad c(\bar{x}, \bar{y}) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad s(\bar{x}, \bar{y}) = \frac{2\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3}
\]

and \(\mu_x\), \(\sigma_x^2\), \(\sigma_{xy}\) are the mean, variance, and covariance of \(\bar{x}\), respectively.

where, \(\bar{x}\) and \(\bar{y}\) are two local windows size \((N \times N)\) of the same spatial location from two images.

\(x_i\) is the \(i\)th pixel of \(\bar{x}\), where \(i = 1, \ldots, N\)

\(\mu_x\) is the mean of \(\bar{x}\)

\(\sigma_x^2\) is the variance of \(\bar{x}\)

\(\sigma_{xy}\) is the covariance of \(\bar{x}\) and \(\bar{y}\)

\(\alpha = \beta = \gamma = 1\) for simplicity

\(C_1\), \(C_2\), and \(C_3\) are constants to prevent the numerator of each term being close to zero (to prevent an unstable measurement).

\(C_1 = (K_1L)^2\), \(C_2 = (K_2L)^2\), \(C_3 = \frac{C_2}{2}\)

\(K_1 = 0.01\), \(K_2 = 0.03\)

\(L = 255\) is the dynamic range of pixel values (for 8-bit per pixel component),
Properties of SSIM index [3, 4],

1. \( \text{SSIM}(x, y) = \text{SSIM}(y, x) \)

2. \( \text{SSIM}(x, y) \leq 1 \)

3. \( \text{SSIM}(x, y) = 1 \) if and only if \( x = y \) (in discrete representations, \( x_i = y_i \) for all \( i = 1, 2, \ldots, N \))

Mean SSIM (MSSIM) index for a single overall quality value of the entire image, is defined as the average of the SSIM map,

\[
\text{MSSIM}(x, y) = \frac{1}{M} \sum_{j=1}^{M} \text{SSIM}(x_j, y_j)
\]

where, \( x_j \) is \( j \) th local window (8x8) of an image. \( M \) is the number of local windows of the image.


\[
\text{SSIM}_{pj} = 0.8 \text{SSIM}^Y_{pj} + 0.1 \text{SSIM}^{Cb}_{pj} + 0.1 \text{SSIM}^{Cr}_{pj}, \quad (2)
\]

\[
Q = \frac{1}{F \cdot M} \sum_{p=1}^{F} \sum_{j=1}^{M} \text{SSIM}_{pj}, \quad (3)
\]

where, \( \text{SSIM}^Y_{pj} \), \( \text{SSIM}^{Cb}_{pj} \), and \( \text{SSIM}^{Cr}_{pj} \) are SSIM index values (1) of Y, Cb, and Cr components of the \( j \) th local window in the \( p \) th video frame. \( F \) is the number of video frames of the video. \( Q \) is the overall quality measure of the entire video sequence.

**LITERATURE REVIEW ON Structural similarity index in video codec design:**

Two papers [7, 8] applied SSIM in JM reference software.

In [7], the author claims that motion estimation method is improved using distortion based on SSIM (MEBSS, motion estimation method based on structural similarity). Both bit rate and coding time are reduced. The simulation is implemented only at \( QP = 10 \), very low quantization parameter value (high quality, high bit rate). Lagrange cost for motion block matching metric is changed from absolute differences to \( (1 - \text{SSIM}) \) with full search motion estimation. SSIM is computed with local block size (4x4) non-overlapped windows.
In [8], the authors propose more techniques to reduce the coding time in motion estimation called FMEBSS, fast MEBSS. SSIM is used in early termination algorithm (with fixed threshold value) and Lagrange cost where absolute differences is changed to $K (1 – \text{SSIM})$. $K$ is the author defined multiplier. The simulation is done at $QP = 10$, 20, and 30. The codec achieves bit rate and coding time reductions.

REFERENCES: