ASSESSMENT OF VIDEO WATERMARKING USING STRUCTURAL METRICS

[1]Divjot Kaur Thind, [2]Sonika Jindal
[1]Student Mtech (Cse), [2]Assistant Professor
SBSSTC, Ferozepur, Punjab
[1]prettythind01@gmail.com, [2]sonikamanoj11@gmail.com

***Abstract*— Digital watermarking was introduced due to rapid advancement of networked multimedia systems. It was developed to enforce copyright technologies for protection of copyright ownership. This technology is first used for still images but recently they have been developed for other multimedia objects such as audio, video etc. In this paper a new digital video watermarking scheme is proposed which combines Discrete wavelet transform (DWT) and Singular Value Decomposition (SVD) in which watermarking is done in the high frequency sub band. Review of many subjective and objective video quality metrics has been done and then various attacks have been applied to calculate the results for these values.**

Keywords—Video watermarking, Structure similarity index (SSIM), MS-SSIM, Discrete wavelet transform (DWT), Singular Value Decomposition (SVD), Gaussian noise, Poisson noise

I. INTRODUCTION

The rapid advancement of digital media (audio, images and video) and their easy replication and distribution has created a need for copyright enforcement techniques. In recent years, digital watermarking has emerged as an effective way to prevent users from violating copyrights. This concept is based on the insertion of information into the data in such a way that the added information is not visible yet robust to any changes made in the watermarked data[1]. In the watermarking algorithms, three factors must be considered:

- Capacity, i.e. the amount of information that can be put into the watermark and recovered without errors;
- Robustness, i.e. the resistance of the watermark to alterations of the original content such as compression, filtering or cropping;
- Visibility, i.e. whether watermark is visible or invisible.

These factors are interdependent so therefore it is very important to consider all these while evaluating the performance. Many benchmarks have been proposed for watermark evaluation [2]. These benchmarks are

- Stirmark: Generic tool for checking robustness of the watermarking algorithms. Performance evaluation is done by using Receiver Operating Characteristics (ROC) that shows relationship between true positive fraction(TPF) and false positive fraction(FPF)
- Checkmark: Extension of stirmark and includes features like new attacks, use more advanced perceptually motivated distortion metrics. It is developed by Shelby Pereira and runs on matlab. Its quality metrics are Watson metrics and weighted PSNR.

- Optimark: Uses multiple trials with different keys and watermark values.
- Certimarkz

II. VIDEO QUALITY ASSESSMENT

There are two methods for quality assessment: Subjective video quality assessment and objective video quality assessment [3]. Subjective quality assessment involves perception of a human being. In this the human rate the visual quality [4]. Objective video quality assessment uses quality metrics to evaluate videos. These metrics are designed mainly on the basis of HVS (Human Visual System) [5].

The benchmark are subjective experiments, where a number of people are asked to watch test clips and to rate their quality. However, subjective experiments require careful setup and are time consuming, hence expensive and often impractical. Due to these advantages simple error measure metrics are used such as PSNR, MSE. These simple conventional metrics are based on the pixel values and neglect various other spatial temporal views.

A. Conventional metrics

1) Mean Square Error(MSE)

It is defined as the difference between values implied by an estimate and the true quality being certificated [6]. MSE is represented as:

$$\mathrm{MSE} = \frac{1}{\mathrm{XY}\left[\sum_{i=1}^{X}\sum_{j=1}^{Y}(c(i,j) - e(i,j))\right]}$$

X and Y are height and width respectively of the image. The c (i, j) is the pixel value of the cover image and e (imp) is the pixel value of the embed image.

2) Peak Signal to noise ratio(PSNR)

It is the ratio between maximum possible power and corrupting noise that affect representation of image. It represents the degradation of the image or reconstruction of an image [6]. It is expressed as a decibel scale. Higher the value of PSNR higher the quality of image. PSNR is represented as:

$$\mathrm{PSNR} = 10\log 10\left(\frac{L * L}{\mathrm{MSE}}\right)$$

3) Signal to noise ratio(SNR)

SNR (Signal to Noise ratio) measures the sensitivity of the imaging. It helps to measures the signal strength relative to the noise. It is calculated by the formula given below:

$$\mathrm{SNR} = 10\log 10\left(\frac{\mathrm{Psignal}}{P_{\mathrm{noise}}}\right)$$

4) Bit error ratio(BER)

BER is the ratio that describes how many bits received
in error over the number of the total bits received.

$$\mathrm{BER} = \frac{P}{(H * W)}$$

B. HVS based metrics

Watermarking introduces special effects in video which have to be properly estimated as suggested in [7]. We have seen mainly two kinds of spatial-temporal impairments.

- Spatial noise: Footprints of the watermark
- Temporal flicker: Visible changes of the watermark pattern between consecutive frames

1) Structural similarity index (SSIM)

The Structural Similarity index (SSIM) is an alternative visual quality metric that is used to measure the similarity between two videos. It is full reference metric [8]. The SSIM has been shown to be more effective at estimating the perceptual quality of images than the PSNR as it considers image degradation as perceived change in structural information. Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close. These dependencies carry important information about the structure of the objects in the visual scene. The SSIM uses luminance, contrast and structure comparison functions to estimate the perceived quality of an image. The result varies from 1.0 to −1.0, where 1 represents perfect quality and −1 very noticeable distortion [9].

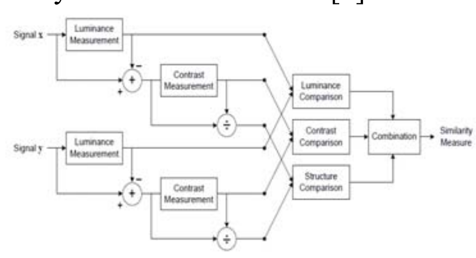

Fig 1: SSIM measurement system

The three comparison stages of the SSIM

a) Luminance Comparison

Assume the two images to be compared are two-dimensional matrices A and B. The mean intensity of each image, μ, is computed as

$$\mathrm{l(A,B)} = \frac{2\mu_A\mu_B + C_1}{\mu_A^2 + \mu_B^2 + C_1}$$

where C_1 is calculated as $C_1 = K_1L_2$ in order to avoid instability when $\mu_A^2 + \mu_B^2$ is close to 0. K_1 is a chosen as a small constant and L is the maximum value any pixel in the image can assume. We use value of 0.01 for K_1 and reports that the SSIM is largely unaffected by changes to this value.

b) Contrast Comparison

The unbiased standard deviation of each image, σ_A and σ_B, are used as an estimate of contrast. The contrast comparison function is given by

$$C(A,B) = \frac{2\sigma_A\sigma_B + C_2}{\sigma_A^2 + \sigma_B^2 + C_2}$$

with $C_2 = (K_2L)^2$ and K_2 chosen as 0.03. It was again reported that the technique is fairly insensitive to small variations in the value of K_2.

c) Structure Comparison

For structure comparison, the mean intensity is subtracted from the pixels in each original image. Each image is then divided by its standard deviation so the two images being compared each have unit standard deviation. For simplicity, we represent A and B in one-dimensional forms a and b. The correlation coefficient σ_{AB} is now defined as

$$\sigma_{\mathrm{AB}} = \frac{1}{XY-1}\sum_{i=1}^{XY}\left(\frac{a_i-\mu_A}{\sigma_A}\right)\left(\frac{b_i-\mu_B}{\sigma_B}\right)$$

and C_3 as a small constant $C_2/2$. The structure comparison function is then given as

$$\mathrm{s(A,B)} = \frac{\sigma_{AB} + C_2/2}{\sigma_A\sigma_B + C_2/2}$$

Now combining all the comparison function that is

$$\mathrm{SSIM(A,B)} = [l(A,B)]^{\alpha}.[c(A,B)]^{\beta}.[s(A,B)]^{\gamma}$$

$$\mathrm{SSIM(A,B)} = \frac{(2\mu_A\mu_B + C_1)(2\sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)}$$

2) Mean Structural similarity index (MSSIM)

This metric is used to measure the overall quality of the video.

$$\mathrm{MSSIM(A,B)} = \frac{1}{\mathrm{M}}\sum_{\mathrm{j}=1}^{\mathrm{M}}\mathrm{SSIM(A_j, B_j)}$$

where A and B are the reference and the distorted images, respectively; A_j and B_j are the image contents at the jth local window; and M is the number of local windows in the image.

3) Multi Scale Structural similarity index(MS SSIM)

SSIM is a single scale metrics where as MS SSIM is a multi scale metrics [10]. MS SSIM is the modification of the SSIM. Fig 2 shows the MS SSIM measurement system in which reference and distorted image is accomplished over multiple scale by iteratively low-pass filtering and down-sampling the signals [11]. Processing in MS-SSIM is simple. The highest scale M is obtained after M-1 iterations. At the jth scale, the contrast comparison and the structure comparison are calculated and the luminance comparison is computed only at Scale M [12] . The overall MS SSIM is obtained after comining all these and represented as:

$$MS\ SSIM(a,b) = l_M(a,b)\prod_{j=1}^{M} c_j\,(a,b)s_j(a,b)$$

Where, l_m is the luminance comparsion and the highest scale used is M=5.

Fig 2: Multi-scale structural similarity measurement.

III. PROPOSED ALGORITHM

This video watermarking scheme uses blue component of RGB colorspace for embedding watermark. Gray watermarks are chosen and are pre-processed such that their size is equal to the size of the DWT sub-band of the video frame. The same watermark is embedded in each frame of the video.

A. *Video Watermark Embedding Process*

Fig 3 shows the basic process of watermark embedding and these steps are explained below

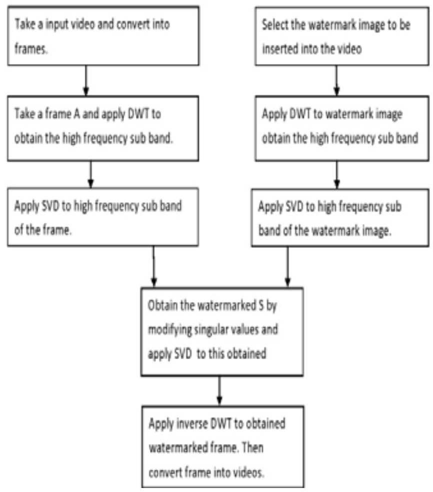

Fig 3: Watermark Embedding Process

1) Video is taken as input which is group of continous frames. An input video is converted into frames. This video is in RGB color space and all the processings are done in the blue component of the RGB color space.

2) Discrete wavelet transform (DWT) is applied to frame A and is decomposed into four sub-bands LL_a,LH_a,HL_a and HH_a.

3) Apply SVD to high frequency sub-band of the original frame.

$$HH_a = U_a^h S_a^h V_a^h$$

4) Take a watermark and then convert into grayscale image. DWT is applied to the watermark image W and decomposed into four sub-bands: $LL_w, LH_w, LH_w$and HH_w.

5) Apply SVD to high frequency sub-band of the watermark.

$$HH_w = U_w^h S_w^h V_w^h$$

6) Modify the singular values of original frame and watermark image and obtain singular value of
watermarked image.

$$S_w^{*h} = S_a^h + k * S_w^h$$

Where k is a scaling factor.

7) Apply SVD on obtained singular value:

$$S_w^{*h} = U_a^{hh} S_a^{hh} V_a^{hh}$$

8) Using DWT to S_a^{hh} , obtain HH_a^*.

$$HH_a^* = DWT(S_a^{hh})$$

9) Apply IDWT to obtain watermarked cover image A* using LL_a,LH_a,$HL_a$and HH_a^*.

B.Video watermarking Extraction Process

Fig 4 shows the basic process of watermark extraction and these steps are explained below

1) Decompose the watermarked image A* into four subbands using DWT: LL_a ,LH_a ,HL_a and HH_a^* .

2) Apply SVD to high frequency sub-band HH*a

$$HH_a^* = U_{wa}^h S_{wa}^h V_{wa}^h$$

3) Decompose the original image A into four sub-bands.

4) Apply SVD on high frequency sub-band of original image, as in embedding process.

$$HH_a = U_a^h S_a^h V_a^h$$

5) Using DWT decompose watermark image W, into four sub-bands and apply SVD to high frequency sub-band as in embedding process:

$$HH_w = U_w^h S_w^h V_w^h$$

6) Extract the singular value of high frequency sub-band watermark image:

$$S = (S_w^h - S_a^h)/k$$

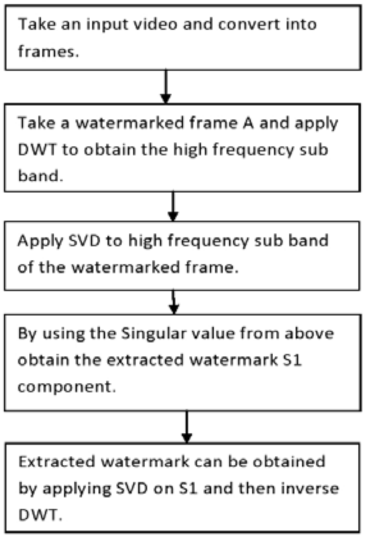

Fig 4: Watermark Extraction Proces

7) Using above S recover the high frequency sub-band of watermark image:

$$HH_w = U_w^h S_w^h V_w^h$$

8) Using LL_w, LH_w, LH_w and HH_w apply IDWT to recover the watermark image W .

IV. RESULTS AND DISCUSSION

The experimental simulation is carried out using matlabR2010b. In this paper we have taken a standard video 'Traffic' as a host video and the watermark is any image. We have taken k as a scaling factor and its value is 0.2. The proposed scheme can perform test on many other videos. The size of the video is 512×512 and the size of watermark is same. The video is in the RGB color space and the processings are done in the blue component of the frame because of its low sensitivity to the human perception. The watermark is converted into grayscale image to reduce its intensity so that algorithm can become more robust.

Fig 5: a) Original First Video Frame b) Watermark c) Watermarked Fisrt frame

Table 1 shows the values of PSNR, BER, MSE, SSIM and MSSIM of the watermarked frames. The value of PSNR is high for this video which means that after embedding the watermark there is very less quality distortion. Since SSIM and MSSIM measures the structure of the images and the maximum value can be 1, which we have obtained this means that after embedding the structure of the video did not changed.

Table 1: The values of different metrics obtained after embedding

PSNR	BER	MSE	SSIM	MSSIM
92.3343	0.0103	0.00003	1	1

Once the watermark embedded, it is ready for transmission. During transmission many processing take place, and these are viewed as attacks such as contrast adjustment and intensity transformation. There are also many other attacks such as geometrical attacks like cropping and rotation, noising attack like Gaussian noise and salt and pepper noise, denoising attacks like average filtering. Table 2 shows all these attacks along with the metrics.

Table 2: Assessment of the metrics using different attacks

Attacks	Image Assessment Metrics				
	PSNR	CF	SSIM	MSSM	MSSSIM
Intensity Transformation	52.6233	0.9997	0.9957	0.9993	0.9997
Frame averaging	51.6003	0.9997	0.9949	0.9992	0.9997
Blur	50.9341	0.9997	0.9947	0.9992	0.9997
Rotation	48.4151	0.9996	0.9916	0.9988	0.9996

Gaussian noise	32.9027	0.9523	0.5846	0.8988	0.9649
Poisson Noise	36.1747	0.9901	0.82285	0.9705	0.9890
Salt and pepper Noise	32.3673	0.9354	0.5296	0.8734	0.9638

V. CONCLUSION

In this paper we have reviewed some new metrices for video watermarking. These metrics are based under assumption that visual perception is highly adapted for extracting structural information from image. We have calculated the values of these metrices on our proposed algorithm. The watermark object has been embedded in each frame of the original video and we obtain high PSNR, SSIM and MS SSIM which means that the quality of the video do not distort. From overall observation it has been established that the proposed scheme yields better imperceptibility and robustness against various attacks which makes the proposedscheme suitable for some application.

REFERENCES

[1] Winkler, Stefan and Gelasca, Elisa Drelie and Ebrahimi, Touradj. “Toward perceptual metrics for video watermark evaluation”, Optical Science and Technology, SPIE's 48th Annual Meeting, pp. 371-378, 2003.

[2] Voloshynovskiy, Sviatoslav and Pereira, Shelby and Iquise, Victor and Pun, Thierry, “Attack modelling: towards a second generation watermarking benchmark”, Signal processing, Elseveir,2001, vol. 81, No. 6, pp.1177-1214.

[3] Seshadrinathan, Kalpana and Soundararajan, Rajiv and Bovik, Alan Conrad and Cormack, Lawrence K, “Study of subjective and objective quality assessment of video” , IEEE, Vol. 19, pp.1427-1441,2010.

[4] Seshadrinathan, Kalpana and Soundararajan, Rajiv and Bovik, Alan C and Cormack, Lawrence K, “A subjective study to evaluate video quality assessment algorithms”, International Society for Optics and Photonics, 2010, pp. 75270H--75270H.

[5] Chikkerur, Shyamprasad and Sundaram, Vijay and Reisslein, Martin and Karam, Lina J, “Objective video quality assessment methods: A classification, review, and performance comparison”, IEEE, vol.57, pp.165-182, 2011.

[6] Singh, Prabhishek and Chadha, RS, “A Survey of Digital Watermarking Techniques, Applications and Attacks”, {International Journal of Engineering and Innovative Technology (IJEIT), 2013, Vol.2, Number. 9.

[7] Winkler, Stefan and Gelasca, Elisa Drelie and Ebrahimi, Touradj, “Perceptual quality assessment for video watermarking”, IEEE,2002, pp. 90-94.

[8] Wang, Zhou and Bovik, Alan C and Sheikh, Hamid R and Simoncelli, Eero P, “Image quality assessment: from error visibility to structural similarity”, IEEE,2004, vol.13, pp.600-612.

[9] Wang, Zhou and Lu, Ligang and Bovik, Alan C, “Video quality assessment based on structural distortion measurement”, Elseveir, 2004, Vol.19, pp.121-132.

[10] Wang, Zhou and Simoncelli, Eero P and Bovik, Alan C, “Multiscale structural similarity for image quality assessment” , IEEE, 2003, Vol. 2, pp.1398-1402.

[11] Li, Chaofeng and Bovik, Alan C, “Content-partitioned structural similarity index for image quality assessment”, Elseveir, 2010, Vol.25, Number.7, pp.517-526.

[12] Fei, Xuan and Xiao, Liang and Sun, Yubao and Wei, Zhihui, “Perceptual image quality assessment based on structural similarity and visual masking” , Elseveir, 2012, Vol.27, pp. 772-783.