<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *14th International Conference on Signal Processing and Communication Systems, ICSPCS 2020, Virtual, Adelaide, Australia, 14 December 2020 through 16 December 2020*.

Citation for the original published paper:

Tran, D N., Zepernick, H-J., Chu, T M. (2020)
Visual Attention Based LSB Data Hiding in 360° Videos
In: Wysocki T.A., Wysocki B.J. (ed.), *2020 14th International Conference on Signal Processing and Communication Systems, ICSPCS 2020 - Proceedings*, 9310051
Institute of Electrical and Electronics Engineers Inc.
<https://doi.org/10.1109/ICSPCS50536.2020.9310051>

N.B. When citing this work, cite the original published paper.

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:bth-21063>

Visual Attention Based LSB Data Hiding in 360° Videos

Dang Ninh Tran, Hans-Jürgen Zepernick, and Thi My Chinh Chu

Blekinge Institute of Technology, SE-371 79 Karlskrona, Sweden

E-mail: {dnt, hjz, cch}@bth.se

Abstract—In this paper, we propose a visual attention based least significant bit (LSB) data hiding method for 360° videos. In particular, information about the relative frequency of pixel access is used to control the amount of secret data to be hidden at the different latitudes of 360° videos. In this way, the typical behavior of users paying more attention to the equator region compared to the north and south poles when viewing 360° videos on a head-mounted display is taken into account. An analytical expression for capacity offered by this method to hide secret data in 360° cover videos is derived. A performance assessment of the visual attention based LSB data hiding method is conducted in terms of the peak signal-to-noise ratio (PSNR) and weighted-to-spherically-uniform PSNR (WS-PSNR) which both quantify the fidelity of a 360° stego-video with reference to the related 360° cover video.

Index Terms—LSB data hiding, 360° videos, visual attention, capacity, PSNR, weighted-to-spherically-uniform PSNR.

I. INTRODUCTION

Immersive digital media such as virtual reality (VR) and augmented reality (AR) together with head-mounted displays (HMDs) have seen rapid development in recent years. In particular, 360° videos offering 360° × 180° viewing directions with respect to longitude and latitude are reported in [1] to account for the majority of VR content. The related so-called weak-interaction VR services offered over wired and wireless networks comprise of IMAX theater, 360° panoramic video streaming, and VR live broadcasting [2]. Apart from stringent network latency constraints such as keeping motion-to-photon latency below 20 ms, networked VR services require substantial bandwidth that may range from 50 Mbps for 360° videos of 4K resolution to above 1 Gbps for 360° videos of 24K resolution. An increased portfolio of digital media including digital twins and holopresence are foreseen to shape sixth generation (6G) mobile communication systems [3].

The large resolutions of 360° videos are essential for providing immersive experiences but may also be used as a resource for concealing large amounts of secret information. Steganography provides covert communication between two parties such that the information exchange cannot be observed by an attacker, i.e., a secret object is hidden in a cover object to constitute a stego-object. In particular, hiding secret information in digital audio, images, and videos exploits the fact that the human auditory system and human visual system are rather insensitive to small changes in digital media. Surveys on the state-of-the-art steganography techniques for these conventional digital media can be found in [4]–[7]. In

order to keep computational load associated with data hiding in digital media low, especially for networked services, least significant bit (LSB) data hiding methods may be used. In the spatial domain, for example, LSB data hiding methods replace the LSB or b rightmost bits of a word or pixel of depth B of the cover object by the secret data. A comprehensive survey on LSB data hiding methods for digital images and videos can be found in [8] providing a classification of hiding data in the spatial, transform, and quantum domain of digital images, and raw and compressed domain of digital videos. As for new digital media such as high-definition images and 360° videos, some LSB data hiding methods are reported in [8]. However, steganography for new digital media is much less developed compared to conventional digital media because the different types of immersive digital media themselves are subject of further development.

Regarding new digital media related to visual stimuli such as 360° videos, additional characteristics related to how humans view scenes on HMDs may be exploited at the different stages of a processing chain in a communication context. In particular, the observation that users statistically focus more on the equator region and the center of the front regions of the shown content compared to the poles has been used for assessing source encoding schemes, and for developing subjective and objective video quality assessment (VQA) methods. In [9], a framework for evaluating 360° video encoding schemes is presented in which viewport based head motion trajectories are extracted. This allows to compare original and encoded videos on the viewport and examine different sphere-to-plane mappings. The work also accounts for the fact that users typically give more attention to the equator region compared to the poles when viewing natural scenes. In particular, average statistics on the relative frequency of pixel accesses versus latitude are provided that were obtained from head motion trajectories over 10 users which viewed 10 different natural 360° video scenes of 10 s duration each. The work reported in [10] proposes subjective and objective methods to assess visual quality loss of 360° videos due to source encoding artifacts. Specifically, a subjective VQA method is presented which takes the typical directions of how users view 360° videos into account. This method reports on the results obtained from a subjective experiment suggesting that viewing directions are normally distributed in the center of the front regions but may also fall into other regions depending on the video content. In addition, two objective VQA methods are proposed to

assess the quality of source encoded 360° videos. One method imposes a weighting on the distortion of pixels subject to their distance to the center of the front region. The other method predicts viewing directions depending on the video content and then allocates weights to the distortions of each pixel based on the predicted viewing direction.

In [11], [12], spherical LSB data hiding in 360° videos and its combination with morphological operations have been proposed. Although the characteristic that viewers typically put more attention to the areas around the equator compared to the poles, the amount of data to be hidden in the selected regions around the poles is kept constant but is not adjusted when moving from the poles to the equator of the 360° video. A fixed number of lines around the poles has been selected instead with each line hiding the same amount of secret data irrespective of the distance to the poles. As for the region around the equator, the insensitiveness of the human visual system to sharp intensity changes in edge regions has been exploited to hide secret data in edges. An edge detector together with morphological operations has been used to increase data hiding capacity.

Motivated by all of the above, in this paper, a visual attention based LSB data hiding method for 360° videos is proposed and its performance is assessed. In particular, statistics on the relative frequency of pixel accesses with respect to the latitude are taken into account for controlling the amount of data to be hidden in different regions of 360° videos. Because the LSB data hiding is controlled by the weights associated with each latitude, the entire field of view is addressed by using the visual attention based LSB data hiding method. In other words, the morphological operations for the equator region used in [12] are not required which simplifies the LSB data hiding procedure. Main contributions of this paper may be summarized as follows:

- A visual attention based LSB data hiding method for 360° videos is proposed.
- The relative frequency of pixel access is used to control the amount of secret data to be hidden depending on the latitude.
- An expression for the capacity provided by this method for hiding secret data in 360° cover videos is derived.
- A performance assessment of the proposed method in terms of video fidelity metrics is provided for a selected 360° cover video and random secret data.

The remainder of this paper is organized as follows. Section II presents the visual attention based LSB data hiding method that accounts for the relative frequency of access of pixels when viewing 360° videos on HMDs. The data hiding capacity, also referred to as capacity for brevity, offered by the proposed visual attention based LSB data hiding method is derived in Section III. Experimental results are provided and discussed in Section IV focusing on the performance of the proposed method based on the mean square error (MSE), peak signal-to-noise ratio (PSNR), weighted MSE (WMSE), and weighted-to-spherically-uniform PSNR (WS-

PSNR). Conclusions are drawn in Section V.

II. VISUAL ATTENTION BASED LSB DATA HIDING

A. Normalized Relative Frequency of Pixel Access

In [9], the relative frequency of pixel accesses according to head motion trajectories for 360° videos viewed on an HMD has been reported. The results are given in terms of average statistics for data obtained from 10 users which viewed 10 360° videos of 10 s duration each. The obtained relative frequency of access suggests that users tend to give higher attention to regions around and near the equator compared to the pole regions. The progression of relative frequency of access versus latitude is provided in graphical form.

Based on selected sample points of the graphical representation given in [9], we have produced an analytical expression $f(\phi)$ of the normalized relative frequency of accesses as function of latitude ϕ using the Matlab curve fitting toolbox. To formulate visual attention based data hiding weights, a normalization is performed with respect to the highest relative frequency of access which results in the following expression:

$$f(\phi) = \begin{cases} \sum_{i=0}^4 a_i \phi^i & \text{for } -90^\circ \leq \phi \leq -40^\circ \\ \sum_{i=0}^6 b_i \phi^i & \text{for } -40^\circ \leq \phi \leq +40^\circ \\ \sum_{i=0}^4 c_i \phi^i & \text{for } +40^\circ \leq \phi \leq +90^\circ \end{cases} \quad (1)$$

The coefficients of the polynomial fit to the sample points of the graphical data are provided in Table I and the resulting progression of normalized relative frequency of access is given in Fig. 1. It can be seen from the figure that the normalized relative frequency peaks at latitudes near $\phi = \pm 30^\circ$ and is slightly lower at the equator with latitude $\phi = 0^\circ$. This characteristic has been explained in [9] as being due to the fact that the viewport related to 360° videos is a projection of the sphere to the plane. In particular, uniform sampling in the plane results in denser samples on the sphere when moving farther away from the equator. Once these peaks are passed, the normalized relative frequency of access decreases with latitudes toward the poles and reaches zero at the south pole $\phi = -90^\circ$ and north pole $\phi = 90^\circ$.

TABLE I
COEFFICIENTS OF THE POLYNOMIAL FITS

i	a_i	b_i	c_i
0	9.570	0.8746	9.570
1	4.851E-01	0	-4.851E-01
2	9.284E-03	6.825E-05	9.284E-03
3	7.927E-05	0	-7.927E-05
4	2.542E-07	3.563E-07	2.542E-07
5	0	0	0
6	0	-3.248E-10	0

Fig. 1. Normalized relative frequency of access versus latitude.

B. Visual Attention Based Data Hiding Weights

The expression $f(\phi)$ in (1) can readily be applied to specify visual attention based data hiding weights as follows. As users tend to give higher attention to the area near the equator, modification of the structure of a scene due to data hiding should be kept small near the equator while increasing amounts of data may be hidden in the viewing regions closer to the poles. The data hiding weights $w(\phi)$ may therefore be simply formulated as complement of the normalized relative frequency $f(\phi)$, i.e.,

$$w(\phi) = 1 - f(\phi) \quad (2)$$

The progression of the visual attention based data hiding weight $w(\phi)$ as a function of latitude ϕ is presented in Fig. 2.

Fig. 2. Visual attention based data hiding weight versus latitude.

An example of a binary map of pixels to be used for data hiding in a frame of a 360° video is shown in Fig. 3. Black pixels indicate positions where data hiding is performed while white pixels are kept free from data hiding. The amount of pixels used for data hiding at each latitude ϕ is determined by the related data hiding weight $w(\phi)$ and placed randomly across the longitude for a given latitude. As such, the progression of $w(\phi)$ shown in Fig. 2 can be observed in the binary map as shown in Fig. 3 when moving from the south pole over the equator to the north pole.

C. LSB Data Hiding

The term least significant bit refers to a bit that has the lowest order in a multiple-bit binary number (see Fig. 4). In

Fig. 3. Binary map of random data hiding positions at each latitude (Black: Pixel used for data hiding; White: Pixel not used for data hiding).

the context of data hiding in digital visual media, the attributes “significant” and “insignificant” also refer to the ability of a bit at bit plane b of a word (or pixel) of bit depth B to be changed with or without perceptual impact on the fidelity or quality of the digital stego-object. A modification of the LSBs of a sequence of words of a cover object by replacing them with the secret data, typically, has little effect on the quality of the obtained stego-object. LSB data hiding may also be extended to higher bit planes, i.e., second LSB, third LSB, and so on. However, the increase of capacity achieved by engaging higher bit planes comes at the expense of quality.

Fig. 4. LSB data hiding in different bit planes of a word (or pixel) of bit depth B ranging from the LSB to the b -th rightmost LSB.

III. CAPACITY

An important metric used for evaluating the performance of data hiding methods is the data hiding capacity or capacity which refers to the number of secret bits that can be embedded in a cover object. For digital videos, capacity can be defined as the average number of secret bits concealed in each video frame which is measured in bits per pixel per frame (bpp/frame). For 360° cover videos with visual attention based LSB data hiding, the cumulative capacity in bpp/frame with respect to increasing latitude for a color or luminance component and a given number of bit planes b can be obtained by integrating (2), i.e.,

$$C(\phi, b) = \frac{b}{180^\circ} \int_{-90^\circ}^{\phi} [1 - f(x)] dx \quad (3)$$

Performing the integration of (3) for the three parts specified in (1) and obtaining the related constants that connect the three integrated parts, the capacity in bpp/frame is obtained as (4).

Fig. 5 shows the progression of the capacity $C(\phi, b)$ of visual attention based LSB data hiding in 360° cover videos as a function of latitude ϕ and different number of bit planes $b =$

$$C(\phi, b) = \frac{b}{180^\circ} \begin{cases} (90^\circ + \phi) - \sum_{i=0}^4 \frac{a_i}{i+1} [\phi^{i+1} - (-90^\circ)^{i+1}] & \text{for } -90^\circ \leq \phi \leq -40^\circ \\ (84.62^\circ + \phi) - \sum_{i=0}^6 \frac{b_i}{i+1} [\phi^{i+1} - (-40^\circ)^{i+1}] & \text{for } -40^\circ \leq \phi \leq +40^\circ \\ (12.35^\circ + \phi) - \sum_{i=0}^4 \frac{c_i}{i+1} [\phi^{i+1} - (40^\circ)^{i+1}] & \text{for } +40^\circ \leq \phi \leq +90^\circ \end{cases} \quad (4)$$

1, 2, ..., 6 bits. It can be observed that significant capacity is cumulated for the latitude progression from -90° to -40° and 40° to 90° while the equator region contributes little capacity for data hiding.

Fig. 5. Capacity $C(\phi, b)$ of visual attention based LSB data hiding in 360° cover videos versus latitude for different number of bit planes.

The total capacity in number of secret bits per frame that can be hidden in a color or luminance component per frame is obtained by multiplying (4) with the resolution $N \times M$ of a given 360° video as

$$C = C(90^\circ, b) \cdot (N \times M) \quad (5)$$

where N and M , respectively, are the width and height of a frame.

Table II shows examples of the total capacity per frame that is provided by popular 360° video resolutions. Depending on the number of bit planes used for data hiding and the selected resolution, total capacity ranges from 1.1 Mbits to 95.3 Mbits. As data hiding degrades the visual quality of the 360° video frames, the wide range of total capacities allows for trading off capacity versus quality such that sufficient data can be hidden per frame while keeping visual quality at a satisfactory level.

The capacity for different video component standards can be obtained as a multiple of the capacity of color or luminance component given in (5). Specifically, for a color encoded format using digital gamma-corrected RGB components, the capacity is obtained as

$$C_{RGB} = 3 \times C \quad (6)$$

while the capacity for the YV12 format used in Section IV is given by

$$C_{YV12} = 1.5 \times C \quad (7)$$

TABLE II
TOTAL CAPACITY PROVIDED BY POPULAR 360° VIDEO RESOLUTIONS

b	$N \times M$ [pixels]	$C(90^\circ, b)$ [bpp/frame]	C [Mbits/frame]
1	7680×3840	0.538	15.9
1	6144×3072	0.538	10.1
1	4096×2048	0.538	4.5
1	3600×1800	0.538	3.5
1	2048×1024	0.538	1.1
3	7680×3840	1.616	47.7
3	6144×3072	1.616	30.5
3	4096×2048	1.616	13.5
3	3600×1800	1.616	10.5
3	2048×1024	1.616	3.3
6	7680×3840	3.232	95.3
6	6144×3072	3.232	61.0
6	4096×2048	3.232	27.1
6	3600×1800	3.232	21
6	2048×1024	3.232	6.8

IV. EXPERIMENTAL RESULTS

A. 360° Cover Video and Secret Data

The 360° video scene “G8Salon” from the publicly available VQA-ODV database [13], [14] has been used here to serve as 360° cover video. A sample frame of this scene in equirectangular projection format is shown in Fig. 6.

Fig. 6. Sample frame of the 360° cover video scene “G8Salon” in equirectangular projection from the sphere to the plane [14].

The 360° video scene “G8Salon” has a resolution of 4096 × 2048 pixels. The 360° cover video produced from this scene is 10 s long and has a frame rate of 30 frames per second (fps). This provides a total of 300 frames that can be used for visual attention based LSB data hiding. Furthermore, the 360° cover video is given in a planar format where each color video frame is represented by separate Y, U, and V components. In particular, the YV12 format was used which is a specific variant of the YUV420 format comprising of a Y frame of resolution $N \times M$ followed by a V frame and a U frame both of resolution $(N/2) \times (M/2)$. It should be mentioned that the YV12 format is more commonly used for many software MPEG codecs compared to other YUV420 formats such as I420 where the luminance component Y is followed by U and V chroma components.

The secret data to be hidden in the 360° cover video has been selected here as random binary bit stream that is produced to match the length related to the targeted capacity. In practical applications, the secret data can be any form of digital media such as digital audio, image, video, or immersive media files.

B. Performance Metrics

- *Peak Signal-to-Noise Ratio*: The PSNR is a fidelity metric that is computationally inexpensive and is defined through the MSE as follows. Consider a 360° video frame of width N and height M , and let $x(i, j, k)$ and $y(i, j, k)$ denote the pixel values of the luminance signal of the k -th original frame and associated k -th processed frame at position (i, j) . Then, the MSE of the k -th pair of original and processed frame is defined as

$$\text{MSE}(k) = \frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} |x(i, j, k) - y(i, j, k)|^2 \quad (8)$$

The PSNR of the k -th pair of frames is obtained with (8) as

$$\text{PSNR}(k) = 10 \log \left(\frac{\text{MAX}^2}{\text{MSE}(k)} \right) \quad (9)$$

where $\text{MAX} = 2^B - 1$ denotes the maximum pixel value for a given bit depth B .

Let the 360° video comprise of a total of K frames. The average MSE and average PSNR, respectively, of an entire video may then be defined as

$$\text{MSE} = \frac{1}{K} \sum_{k=0}^{K-1} \text{MSE}(k) \quad (10)$$

$$\text{PSNR} = 10 \log \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \quad (11)$$

- *Weighted-to-Spherically-Uniform PSNR*: The WS-PSNR [15], [16] accounts for the non-linear relationship and warping effect caused by the mapping between plane and sphere. It measures the distortions of samples in the plane and weights these distortions with the projection area of the related surface on the sphere. In particular, the weighted mean squared error (WMSE) of the k -th pair of original and processed frame of

a 360° video can be formulated using the definition given in [15] as

$$\text{WMSE}(k) = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \{ [x(i, j, k) - y(i, j, k)]^2 w(i, j) \}}{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} w(i, j)} \quad (12)$$

where the weights for the equirectangular projection used in this paper are given as

$$w(i, j)_{erp} = \cos \left[\frac{(j + 0.5 - M/2)\pi}{M} \right] \quad \text{and} \quad M = \frac{N}{2} \quad (13)$$

Given (12), the WS-PSNR is defined as

$$\text{WS-PSNR}(k) = 10 \log \left(\frac{\text{MAX}^2}{\text{WMSE}(k)} \right) \quad (14)$$

Similar as for the average MSE and average PSNR in (10) and (11), respectively, the WMSE and WS-PSNR of a 360° video comprising of a total of K frames may be defined as

$$\text{WMSE} = \frac{1}{K} \sum_{k=0}^{K-1} \text{WMSE}(k) \quad (15)$$

$$\text{WS-PSNR} = 10 \log \left(\frac{\text{MAX}^2}{\text{WMSE}} \right) \quad (16)$$

- *Cumulative Performance Metrics*: The MSE and WSME given in (10) and (15), respectively, may be expressed as functions of latitude ϕ , i.e., $\text{MSE}(\phi)$ and $\text{WSME}(\phi)$ focusing on the error due to LSB data hiding at latitude ϕ of the frames of a 360° cover video. The cumulative MSE and cumulative WSME quantifying accumulated errors with respect to increasing latitude can then be defined as

$$\text{C-MSE}(\phi) = \sum_{l=-90^\circ}^{\phi} \text{MSE}(l) \quad (17)$$

$$\text{C-WMSE}(\phi) = \sum_{l=-90^\circ}^{\phi} \text{WMSE}(l) \quad (18)$$

The PSNR and WS-PSNR of the related cumulated errors, respectively, are obtained with (17) and (18) as

$$\text{PSNR}(\phi) = 10 \log \left(\frac{\text{MAX}^2}{\text{C-MSE}(\phi)} \right) \quad (19)$$

$$\text{WS-PSNR}(\phi) = 10 \log \left(\frac{\text{MAX}^2}{\text{C-WMSE}(\phi)} \right) \quad (20)$$

Fig. 7 shows the $\text{MSE}(\phi)$, $\text{C-MSE}(\phi)$, and $\text{PSNR}(\phi)$ of the 360° stego-video with reference to the 360° cover video “G8Salon” specified in Section IV-A for random LSB data hiding in $b = 1$ bit plane. It can be observed that the mean square error $\text{MSE}(\phi)$ follows the progression of the data hiding weights $w(\phi)$ presented in Fig. 2. Accordingly, the mean square errors are the largest at latitudes near the poles and become smaller for latitudes towards the equator. As a result, the cumulative MSE $\text{C-MSE}(\phi)$ increases significantly for latitudes from -90° to -40° and again from 40° to 90° while

(a) MSE as a function of latitude

(b) Cumulative MSE

(c) PSNR(ϕ) obtained from the cumulative MSE

(a) WMSE as a function of latitude

(b) Cumulative WMSE

(c) WS-PSNR(ϕ) obtained from the cumulative WMSE

Fig. 7. Fidelity in terms of the MSE as a function of latitude, C-MSE(ϕ), and PSNR(ϕ) of the 360° stego-video with reference to the 360° cover video “G8Salon” for random LSB data hiding using $b = 1$ bit plane.

Fig. 8. Fidelity in terms of the WMSE as a function of latitude, C-WMSE(ϕ), and WS-PSNR(ϕ) of the 360° stego-video with reference to the 360° cover video “G8Salon” for random LSB data hiding using $b = 1$ bit plane.

little distortions are cumulated for latitudes from -40° to 40° . The PSNR(ϕ) obtained from the cumulative MSE decreases steadily as more latitudes are used for LSB data hiding. Due to the characteristics of the logarithmic function, PSNR(ϕ) decreases steeply for latitudes from -90° to around -40° while further increase of latitudes used for LSB data hiding does not significantly decrease PSNR(ϕ) further in terms of dB.

Similarly, Fig. 8 shows the WMSE(ϕ), C-WMSE(ϕ), and WS-PSNR(ϕ) of the 360° stego-video with reference to the 360° cover video “G8Salon” for random LSB data hiding in $b = 1$ bit plane. Because of the weights for the equirectangular projection in (13), the weighted mean square errors WMSE(ϕ) at the poles are reduced. The cumulative WMSE C-WMSE(ϕ) is therefore kept lower compared to the cumulative MSE C-MSE(ϕ). As a result, the WS-PSNR(ϕ) obtained from the cumulative WMSE is higher compared to the PSNR(ϕ).

C. Cumulative MSE and related PSNR

Let us now examine the impact of engaging higher bit planes into visual attention based LSB data hiding on the MSE(ϕ) and PSNR(ϕ).

Figs. 9(a)-(b) show the cumulative MSE C-MSE(ϕ) and related PSNR(ϕ) for the 360° stego-video of length 300 frames obtained by embedding random secret data into the 360° cover video “G8Salon”. The number of bit planes used for LSB data hiding is varied from $b = 1$ bit to $b = 6$ bits. It can be observed that C-MSE(ϕ) becomes more severe when $b \geq 4$ bits are used for LSB data hiding and is significant for $b = 6$ bits. The visual attention based weighting with respect to increasing latitude ϕ from south pole ($\phi = -90^\circ$) to north pole ($\phi = 90^\circ$) is clearly seen in the progression of C-MSE(ϕ). Specifically, a minor increase of C-MSE(ϕ) is observed in the mid-region ($-40^\circ \leq \phi \leq 40^\circ$) around the equator where viewers pay highest attention to. A steep

(a) Cumulative MSE

(b) PSNR(ϕ) obtained from the cumulative MSE

Fig. 9. Impact of the number of bit planes on the cumulative MSE with respect to latitude and related PSNR(ϕ) of the 360° stego-video with reference to the 360° cover video “G8Salon”.

increase of $C\text{-MSE}(\phi)$ is induced by the visual attention based LSB data hiding in the pole regions $-90^\circ \leq \phi \leq -40^\circ$ and $40^\circ \leq \phi \leq 90^\circ$ which is affordable as users pay less attention to these regions.

The related PSNR(ϕ) shows that the fidelity decreases fast with the increase of the number of bit planes used for LSB data hiding. It should be mentioned that the symmetry of the progression obtained for the $C\text{-MSE}(\phi)$ with respect to the equator $\phi = 0^\circ$ does not carry over to the PSNR(ϕ) which takes the logarithm on the inverse of the $C\text{-MSE}(\phi)$ (see (19)).

D. Cumulative WMSE and related WS-PSNR

Figs. 10(a)-(b) depict the $C\text{-WMSE}(\phi)$ and WS-PSNR(ϕ) obtained for the same 360° stego-video as in Section IV-C.

The general progressions of $C\text{-WMSE}(\phi)$ and WS-PSNR(ϕ) and the impact of the number of bit planes used for visual attention based LSB data hiding on the fidelity is similar to those for $C\text{-MSE}(\phi)$ and PSNR(ϕ). However, compared to the $C\text{-MSE}(\phi)$ for a given number of bit planes, the $C\text{-WMSE}(\phi)$ is kept significantly lower for all latitudes. Furthermore, the $C\text{-WMSE}(\phi)$ is more flat at the pole regions because of the additional weighting (13) used to calculate the WMSE in (12). It is recalled that this weighting has been introduced to account for the warping effect induced by the mapping between plane and sphere. For example, distortions along the latitudes representing the south pole or north pole in the equirectangular projection format are mapped to a single

(a) Cumulative WMSE

(b) WS-PSNR(ϕ) obtained from the cumulative WMSE

Fig. 10. Impact of the number of bit planes on the cumulative WMSE with respect to latitude and related WS-PSNR(ϕ) of the 360° stego-video with reference to the 360° cover video “G8Salon”.

point at the poles on the sphere. As such, the $C\text{-WMSE}(\phi)$ takes the warping effect and related redundancy caused by the mapping between plane and sphere into account. The fact that $C\text{-WMSE}(\phi)$ is kept lower compared to $C\text{-MSE}(\phi)$ in turn results in WS-PSNR(ϕ) being kept higher than PSNR(ϕ).

Table III provides a comparison of the PSNR(ϕ) and WS-PSNR(ϕ) at the poles, i.e., $\phi = -90^\circ$ and $\phi = 90^\circ$. It may be conjectured that PSNR(ϕ) provides a more pessimistic assessment of the fidelity of the 360° stego-video.

TABLE III
COMPARISON OF THE CUMULATIVE FIDELITY AT THE POLES OBTAINED BY DIFFERENT METRICS FOR DIFFERENT NUMBER OF BIT PLANES

b	South pole ($\phi = -90^\circ$)		North pole ($\phi = 90^\circ$)	
	PSNR(ϕ)	WS-PSNR(ϕ)	PSNR(ϕ)	WS-PSNR(ϕ)
1	77.27	99.47	53.84	55.66
2	70.29	92.52	46.84	48.64
3	63.70	85.91	40.56	42.38
4	56.64	78.95	34.51	36.37
5	54.53	76.63	28.74	30.48
6	46.61	68.79	22.72	24.47

V. CONCLUSIONS

In this paper, we have proposed and assessed the performance of a visual attention based LSB data hiding method for 360° videos. Specifically, the user behavior of focusing more on the equator region when viewing 360° videos on HMDs has been taken into account. Visual attention based data hiding weights have been used that control the amount of secret data to be hidden at different latitudes. Accordingly, a larger amount of secret data is hidden in the pole regions while decreasing amounts of secret data is hidden for latitudes that are closer to the equator. An analytical expression for the capacity of the proposed method has been derived. The expression has been used to assess the impact of the number of bit planes and 360° video resolution on capacity. A performance assessment of the visual attention based LSB data hiding method has been conducted using the PSNR and WS-PSNR. The numerical results suggest that the PSNR based evaluation provides a more pessimistic assessment of the fidelity of a 360° stego-video compared to WS-PSNR. This is because WS-PSNR accounts for the warping effect caused by the projection between plane and sphere.

In view of the results presented in this paper, promising directions for future work include the following. Because PSNR and WS-PSNR are based on the differences between pixel intensities, these metrics measure the fidelity of the 360° stego-video. Subjective experiments are therefore needed to produce a ground truth on how humans perceive the quality of 360° stego-videos generated by visual attention based LSB data hiding with different number of bit planes and resolutions. Apart from the frequency of pixel access with respect to latitude, a more general experimental design could also include the frequency of pixel access with respect to longitude. In addition, given that an HMD provides user's only with a limited viewport rather than the entire sphere at a given time, more advanced visual attention models and adaptive LSB data hiding with respect to the viewport dynamics may be considered in future work. On this basis, objective perceptual quality metrics may be developed that are able to predict the quality of the visual attention based LSB data hiding method or other data hiding methods on the quality as perceived by humans.

ACKNOWLEDGMENTS

This work was supported in part by The Knowledge Foundation, Sweden, through the ViaTech project (Contract 20170056). Dang Ninh Tran has been supported by a VIED scholarship awarded by the Vietnamese Government.

REFERENCES

- [1] Huawei iLab, *VR Big Data Report*, Huawei Technologies Co., Ltd., P. R. China, 2017.
- [2] —, *Cloud VR Network Solution White Paper*, Huawei Technologies Co., Ltd., P. R. China, 2018.
- [3] S. Yrjölä, P. Ahokangas, and M. Matinmikko-Blue (Eds.), *White Paper on Business of 6G. (6G Research Visions, No. 3)*, University of Oulu, Finland, Jun. 2020.
- [4] F. Djebbar, B. Ayad, K. A. Meraim, and H. Hamam, "Comparative Study of Digital Audio Steganography Techniques," *EURASIP J. on Audio, Speech, and Music Process.*, vol. 2012, no. 1, pp. 1–16, Oct. 2012.
- [5] A. Cheddad, J. Condell, K. Curran, and P. M. Kevitt, "Digital Image Steganography: Survey and Analysis of Current Methods," *Signal Process.*, vol. 90, no. 2, pp. 727–752, Mar. 2010.
- [6] I. J. Kadhimi, P. Premaratne, P. J. Vial, and B. Halloran, "Comprehensive Survey of Image Steganography: Techniques, Evaluations, and Trends in Future Research," *Neurocomputing*, vol. 335, pp. 299–326, Mar. 2019.
- [7] Y. Liu, S. Liu, Y. Wang, H. Zhao, and S. Liu, "Video Steganography: A Review," *Neurocomputing*, vol. 335, pp. 238–250, Nov. 2019.
- [8] D. N. Tran, *On LSB Data Hiding in New Digital Media*, Licentiate Dissertation, Blekinge Institute of Technology, Karlskrona, Sweden, May 2020.
- [9] M. Yu, H. Laksman, and B. Girod, "Framework to Evaluate Omnidirectional Video Coding Schemes," in *Proc. IEEE Int. Symp. on Mixed and Augmented Reality*, Fukuoka, Japan, Sep./Oct. 2015, pp. 31–36.
- [10] M. Xu, C. Li, Z. Chen, Z. Wang, and Z. Guan, "Assessing Visual Quality of Omnidirectional Videos," *IEEE Trans. Circuits and Systems for Video Technol.*, vol. 29, no. 12, pp. 3516–3530, Dec. 2019.
- [11] D. N. Tran and H.-J. Zepernick, "Data Hiding in 360-Degree Videos With Equirectangular Projection," in *Proc. Int. Conf. on Advanced Technologies for Commun.*, Hanoi, Vietnam, Oct. 2019, pp. 56–62.
- [12] —, "Spherical LSB Data Hiding in 360° Videos Using Morphological Operations," in *Proc. Int. Conf. on Signal Processing and Commun. Systems*, Gold Coast, Australia, Dec. 2019, pp. 573–582.
- [13] L. Chen, X. Mai, D. Xinzhe, and W. Zulin, "Bridge the Gap Between VQA and Human Behavior on Omnidirectional Video: A Large-Scale Dataset and a Deep Learning Model," in *Proc. ACM Int. Conf. on Multimedia*, Seoul, Republic of Korea, Oct. 2018, pp. 932–940.
- [14] *VQA-ODV*, Beihang University, School of Electronic and Information Engineering, Beijing, China, 2018 (accessed June 1, 2020). [Online]. Available: <https://github.com/Archer-Tatsu/VQA-ODV>
- [15] Y. Sun, A. Lu, and L. Yu, "Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1408–1412, Sep. 2017.
- [16] *360tools: Projection and Quality Evaluation Tools for VR Video Compression Exploration Experiments*, Samsung, Republic of Korea, 2016, (accessed May 23, 2019). [Online]. Available: <https://github.com/Samsung/360tools>