An Analysis on Pixel Redundancy Structure in Equirectangular Images

Vazquez, I.

Boise State University

1910 W University Dr.

USA 83725, Boise, ID

ikervazquezlopez@u.boisestate.edu

Cutchin, S.

Boise State University

1910 W University Dr.

USA 83725, Boise, ID

stevencutchin@u.boisestate.edu

ABSTRACT

$360°$ photogrammetry captures the surrounding light from a central point. To process and transmit these types of images over the network to the end user, the most common approach is to project them onto a 2D image using the equirectangular projection to generate a $360°$ image. However, this projection introduces redundancy into the image, increasing storage and transmission requirements. To address this problem, the standard approach is to use compression algorithms, such as JPEG or PNG, but they do not take full advantage of the visual redundancy produced by the equirectangular projection. In this study of the 360SP dataset (a collection of Google Street View images), we analyze the redundancy in equirectangular images and show how it is structured across the image. Outcomes from our study will support the developing of spherical compression algorithms, improving the immersive experience of Virtual Reality users by reducing loading times and increasing the perceptual image quality.

Keywords
Image compression, Equirectangular projection, Virtual Reality,

1 INTRODUCTION

Virtual Reality (VR) provides real-world scenarios focusing on maximizing the immersive sensation. Images and videos shown in VR devices are known as panoramic, $360°$ or omnidirectional images. Gaming, art, photography and social media takes advantage of the immersive properties VR provides and has been used in many applications [Gooed, Faced].

$360°$ photogrammetry captures the surrounding light from a central point. To process and transmit this type of image over the network to the end user, the most common approach is to project them onto a 2D image using the equirectangular projection to generate a $360°$ image. This permits easy image storage and visualization but introduces a major problem. Because the projection reduces the dimensionality from 3D to 2D, it generates topological visual alterations related to the projection stretching samples over multiple pixels and introducing redundant data.

To the best of our knowledge, there is no study about how the redundancy is structured in equirectangular images, nor how much each pixel contribute to the overall image quality. In order to provide a good immersive experience by reducing image loading times and improve the streaming quality, it is important to determine how the redundant data is structured in equirectangular images. With the current trend of resolution increase in visualization devices, images can be adapted accordingly. If we increase the image resolution, the number of redundant pixels will also increase and image file size will be needlessly increased. Redundant pixels do not contribute to the image quality as much as unique pixels, and their impact on the user's immersive experience is often minimal.

In this paper, we present the results of a study we conducted on how perceptual redundancy is structured in equirectangular images. The redundancy is created by how the equirectangular projection stretches the captured samples from the $360°$ field of view. We determine what is the contribution of each of the projected image pixels to the image's overall perceptual quality. Because of the nature of equirectangular projection the portions of the image at angles farther from the horizon (top and bottom rows respectively) contribute less to the overall image quality than the portions nearest the image horizon. Since equirectangular images capture a $360°$ field of view, some areas of the image are stretched because the scene is projected creating low

gradient textures (top and bottom rows) while other areas remain unstretched (middle rows). The contribution of this paper is an analysis, using different metrics (row and region based), of how the redundancy is structured in equirectangular images and how each portion in the image contributes differently to the overall image quality.

2 RELATED WORK

360° photogrammetry captures the surrounding light using a camera located at the central point in a scene. There are multiple ways to capture these type of photos, including multi-shot camera rigs, single camera shot and single camera shot with catadioptric lenses. By applying warping and/or stitching techniques to the captures, we can build a 360° photo in which each pixel represents a small portion of the full scene [Sze10]. Conceptually, the photo is a colored spherical point cloud in a 3D space. The density of this point cloud determines the resolution of the capture and therefore the overall quality of the photo. Using higher resolution sensors in cameras results in a direct visual quality increase. However, the memory requirement for 360° photos that provide a good visual quality is high.

For efficient storage and transmission of 360° photos, the photo sphere must be projected from 3D space to 2D to generate a 360° image. This process produces visual alterations in the projected image yielding redundant data. The most common projections are cylindrical, equirectangular and cube-map projections. As *Azevedo et al.* [Aze19] stated, the capturing, processing and delivery of 360° content, introduces visual alterations which do not fully keep perspective projection properties. The study reviewed the most common types of visual alterations, such as geometrical and data alterations, and identified their main causes.

Visual alterations generate redundant data that can be compressed to reduce the storage and transmission requirements because they do not add any improvement to image quality. Despite the good performance of standard codecs for regular images, using them to compress 360° images does not take full advantage of the 360° image's spherical properties. *Sun et al.* [Sun18] conducted a subjective image quality assessment study on how common image codecs influence the 360° image quality. And it has been shown that codecs for 2D images can be adapted to the 360° images by exploiting projection properties [DS16].

Downsampling specific regions in equirectangular images was proposed in some works. Since top and bottom rows of the equirectangular mapping are stretched, Budagavi *et al.* [Bud15] proposed a Gaussian smoothing filtering over those regions. The key point in that work was the exploiting of the characteristics of the projection and give more importance to equatorial regions. Using the same idea, Youvalari *et al.* [You16], split the image in several strips and each one is downsampled differently based on the latitude they are located. Lee *et al.* [Lee17] used a similar approach but in a pixel level, where each image row is individually downsampled and produces a rhomboid shape image. This studies remove most of the redundant data produced by the equirectangular projection, however, they do not fully consider the structure of visual alterations produced in the projected images.

The compression efficiency of the 360° images differs based on the used projection [Jam19, Yu15]. Some studies considered compression approaches by just re-projecting the images and discovered that compression performance is content dependent. Boyce *et al.* [Boy17], detected regions where textures were highly detailed and rotated the panoramas in the spherical domain to relocate those textures close to the equator. In this manner, regions with low gradients will locate in top and bottom rows where the stretching is higher and benefits the compression with standard compression algorithms. Sun *et al.* [Sun18], proposed the addition of deep learning to the process and by designing a Convolutional Neural Network measured the relation between the visual content and compression rate for different rotations.

The most common metrics to measure the effectiveness of the proposed compression approaches are Peak Signal to Noise Ratio (PSNR) and Structural Similarity Metric (SSIM) [Wan04]. These two metrics were designed to approximate the human perception of regular 2D images, however, they may not provide accurate measurements for 360° images. Since the spherical captures are projected onto a 2D plane, redundant pixels emerge and therefore these metrics are biased by the repeated pixels. To address this problem, the WS-PSNR [Yu15] and S-SSIM [Che18] were proposed, where the results of PSNR and SSIM are weighted, giving more weight to elements close to the horizon.

There exists multiple benchmarks for studying 360° images in different areas. Datasets, such as Abreu *et al.* [DA17], Sitzman et al. [Sit18], Gutierrez et al. [Gut18], were presented to study how the users visualize 360° images by tracking their attention, head movement and eye movement. For 360° image visual quality assessment, the most popular dataset are Huang et al. [Hua18], CVIQD [Sun17], CVIQD2018 [Sun18] and OIQA [Dua18]. SUN360 [oTed] and 3D60 [oITed] (which is a mix of the SunCG [Soned] and SceneNet [McCed] datasets) and 360SP [Cha18] are interesting benchmarks too. 360SP contain outdoor images intended to locate the sun position while 3D60 is a synthetic dataset of indoor images with their respective depth maps. Even though these benchmarks were not

intended to analyze the visual alterations of 360° images, they are a good source for analyzing the redundancy structure in equirectangular images.

3 METHODOLOGY

The main idea of this study is to assess how much individual pixels contribute to the overall image quality based on their location in an equirectangular image. Due to the projected visual alteration when projecting a 3D sphere onto a 2D image plane, some of the samples are stretched across multiple pixels and redundant pixels emerge. Those redundant pixels do not contribute as much to the image quality as non-redundant pixels and add extraneous data that is needlessly stored and/or transmitted. Knowing how this data is structured across equirectangular images is key to designing efficient 360° image focused compression algorithms.

In this work, we first analyzed the quality impact of each equirectangular image row. In each row, we remove multiple sets of contiguous pixels and interpolate gaps to demonstrate the ease of their reconstruction. By computing the resulting row quality, this approach shows the redundancy behavior and provides a visual clue to the number of extra pixels in the image. Full redundancy structure, however, is not revealed with a row based approach and so we use region focused approaches for a better revelation of the underlying structure. Instead of computing quality metrics over entire rows, for each pixel in the image, we focus on its neighbouring region and compute its quality after the region interpolation. In both cases, when we compute the quality metric, highly redundant regions yield high quality measures even when pixels are removed while the quality drops in low redundancy regions when pixels are removed.

With these approaches, the study shows some bias under certain image conditions (such as smooth textures being in the same region). To ensure a robust analysis, we discuss this problem and possible solutions in Section 4.

3.1 Row impact

In equirectangular images, portions of the image at angles farther from the horizon have higher pixel redundancy than at the horizon line due to the projection properties. To test how this redundancy is structured along the image rows, we remove sets of pixels in an increasing number, interpolate them and compute a quality metric. This process reveals the behavior of redundancy in equirectangular images.

We extracted each row in the image and processed them individually to compute their contribution to the image quality. As seen in Figure 1, from each row, we removed multiple sets of contiguous pixels and kept some

 Reference pixel.

 Interpolated pixel.

Interpolated pixels: 1

Interpolated pixels: 2
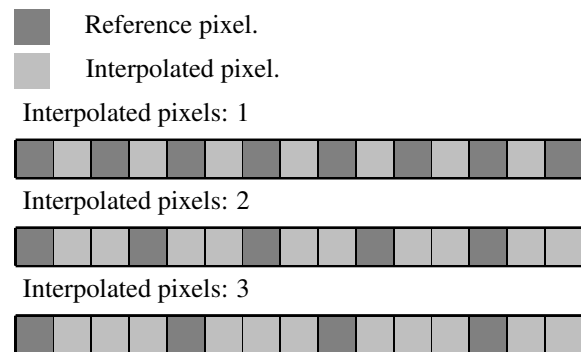

Interpolated pixels: 3

Figure 1: Image row pixel removal and interpolation in an increasing manner.

of the original image pixels as reference. We increase the size of the removed sets in an iterative manner to decrease the number of reference pixels and increase the image error. Then, we reconstruct the missing pixels with the linear interpolation using the reference ones. Since the linear interpolation is a basic interpolation technique and leads to poor results for image reconstruction, it shows how easy it is to reconstruct some regions in equirectangular images. Our thesis is that if, after reconstructing the row with a very basic technique, the resulting row quality is high, the redundancy in that row would be high and therefore most of the pixels in it do not have a high impact on the image quality. Yet, if the resulting quality is low, the redundancy in that row is minimal and most of the pixels in that row would be relevant to the overall equirectangular image quality.

When computing the quality metric, our focus is on the metric behavior and not on its specific value. We wanted to analyze the redundancy behavior in equirectangular images as we move from the top to the bottom rows. Because we processed each row individually, the rest of the image remains untouched.

Computing the row based quality metric over the entire image is unnecessary because it will produce values close to 0 and, based on the image size, the arithmetic precision was seen to affect the results. To address this problem, we computed the row based quality metric only over the modified row and not over the full image. This approach does not have an impact on the study because we are analyzing the error behavior and not the error value. The behavior will be the same even if we compute the row based quality metric of the entire image or just a single row, the only difference will be the resulting error values.

As a result of this process, Algorithm 1, we produce a 2D matrix containing the impact of each row with different removed pixel set sizes. The first dimension of the matrix represents the rows while the second dimension represents the size of removed pixel sets.

Algorithm 1: Row impact on image quality

Function *Row_Impact(img) : matrix* **is**
   q = matrix[img.rows, MAX_SET_SIZE];
   foreach *row* ∈ [0..*img.rows*] **do**
      foreach *ss* ∈ [1..*MAX_SET_SIZE*] **do**
         r = remove_pixels(img[row], ss);
         r' = interpolate_pixels(r, ss);
         q[row, ss] = quality(img[row], r');
      end
   end
   return *q*
end

3.1.1 Metrics

The quality measurement of the reconstructed rows can be computed using different metrics. In this work, for standard image comparison, we use PSNR and SSIM to compare the equirectangular images. However, these metrics do not take into account the properties of 360° images and therefore, the visual from projecting the 3D sphere onto a 2D plane may affect the results.

$$w(r) = cos\frac{\left(r + 0.5 - \frac{N}{2}\right)\pi}{N} \qquad (1)$$

$$WMSE(r) = \frac{\sum_{i=0}^{cols}(x_i - y_i)^2 \cdot w(r)}{\sum_{j=0}^{rows} w(j)} \qquad (2)$$

$$WPSNR(r) = 10log\left(\frac{MAX_I^2}{WMSE(r)}\right) \qquad (3)$$

where:

r = image row.
N = image height.
x = original image.
y = reconstructed image.
MAX_I = maximum image intensity value.

In addition to the standard metrics, we use their weighted versions: W-PSNR and S-SSIM. These two metrics weigh the resulting quality using a row based cosine weight, they lower values from top and bottom rows while they rise values in the middle rows. We slightly modified them to only compute the values from single rows instead of over the full image (Equation 3 for $WPSNR_{row}$ and Equation 5 for S_SSIM_{row}).

$$SSIM(r) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \qquad (4)$$

where:

μ = image row mean.
σ_{xy} = x and y image row covariance.
σ^2 = image row variance.
c_1 = $(0.01 \cdot MAX_I)^2$
c_2 = $(0.03 \cdot MAX_I)^2$

$$S\_SSIM(r) = \frac{\sum_{i=0}^{cols} SSIM(r) \cdot w(r)}{\sum_{j=0}^{rows} w(j)} \qquad (5)$$

3.2 Pixel impact

Studying how the rows impact the equirectangular image quality provides an overall idea about how pixels are relevant in the projected image. A limitation to this approach is that it does not provide information about pixel-wise impact or how the redundancy is structured throughout equirectangular images. Because the equirectangular projection maps the meridians to straight lines the rows are stretched and redundancy emerges perpendicular to the meridians. In order to conduct a study about how each pixel impacts the image quality, we use the ROAD [Gar05] metric and localized SSIM.

In the pixel impact analysis, we do not use the PSNR metric because it is resolution dependant. PSNR requires the computation of the Mean Squared Error, and in small regions, a small difference highly impacted the resulting perceptual quality, therefore, the variance of PSNR results over multiple image is too large to be useful when averaging them.

3.2.1 ROAD metric

The ROAD metric is defined for each pixel (x, y) as Equation 6, where r is the sorted difference of a pixel respect it's neighborhood ones in an ascending order, and $2 \leq m \leq 7$. This metric was presented as a measurement to detect impulse errors when transmitting images over the network. It is a metric that measures the gradient of a pixel respect to its most similar neighborhood ones in a single image. Since the projection generates visual effects in a similar manner on all equirectangular images and stretches pixels across the rows, this metric provides insights about regions with low gradients. Therefore, we only computed the ROAD metric over the original equirectangular image and not over the interpolated image.

$$ROAD_m(x, y) = \sum_{i=0}^{m} r_i(x, y) \qquad (6)$$

There are two possible reasons for the low gradient regions: smooth textures and projection's visual alterations. With the ROAD metric, it is not possible to discern, within a single image, whether or not the low gradient regions belong to smooth textures or to visual alterations. However, by computing the ROAD metric

over a number of images and averaging the results, we introduce variability into the equation and therefore we can eliminate smooth texture effects and only the projection's visual alterations remain.

3.2.2 Localized SSIM

In the standard approach, when computing the SSIM metric, the entire image is used to generate one single value. This value represents the quality of a reconstructed image with respect to the original one. Because in our case we want to measure the impact of a pixel in its local region after reconstructing the image, we created a modified SSIM to only be computed in local regions.

For each pixel in the image, we cropped the 3×3 neighborhood from both images (original and reconstructed) and treated them as full images to compute the SSIM. This operation reveals how much that region impacts in the overall image quality. When generating the results over the entire image, we reveal the redundancy structure in the image, as explained in Section 5.

4 EVALUATION

One of the main problems we found for this work was the configuration of the captured 360° photos. Usually, for better visualization, the horizon line of the scene is mapped to the equirectangular image's middle rows. This orientation provides a good overall idea about how the image will look in a 360° viewer when the content is projected onto a sphere. Because of that, the sky and ground tend to be located in the top and bottom rows. Those regions tend to be smooth textures with very little change and therefore, the redundancy structure analysis would be biased, resulting in highly redundant regions when analyzing them. In order to address this, as stated in Section 3.2, we introduce variability by vertically rotating (pitch) the 360° images in the dataset using 22 degree intervals. To compute the vertical rotation, we transform the equirectangular image to the 3D sphere, rotate it in the x axis and project it back to the 2D image using the equirectangular projection. This operation moves all regions in the original images to different locations in the image, which ensures that smooth regions will be processed at different locations, ensuring a robust redundancy analysis.

In our analysis, variability is important and therefore, we require a large non-synthetic 360° image dataset. Synthetic datasets tend to contain large amounts of smooth textures which will bias our results by not providing enough variability. The 360SP dataset is the perfect fit for our needs. It is composed of $\approx$15000 outdoor images collected from Google Street view at 3328×1664 resolution. After our rotations, the number of images we generated from the dataset was ≈ 170000, which makes a large enough dataset that provides the needed variability to our analysis.
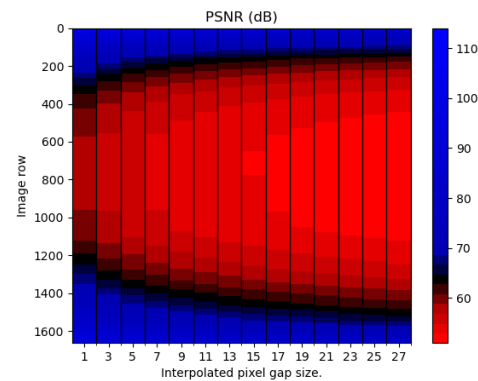

Figure 2: PSNR analysis results.

5 RESULTS

In this section, we show and discuss the results of computing the metrics presented in Section 4. It is important to keep in mind that our results were not intended to compare any methods specifically for their effectiveness. Our goal with this analysis was to study the metric behavior and not the actual values. We ignored the results in terms of their value and focused on how each metric behaved across the equirectangular image. By studying how the metric changed for different pixel gap sizes in each row and in each pixel's neighboring region, we revealed the redundancy structure in equirectangular images.

In Figure 2, we can see that the perceptual impact is minimal in top and bottom rows. It is interesting to point out that PSNR does not seem to care that much about the structure of the equirectangular image, it is sensitive to noise in highly redundant regions. As it is shown in Figure 2 at image rows $[200, 400]$, a small change in a smooth region highly impacts the image's PSNR.

SSIM metric shows more robust results than PSNR for highly redundant regions. In Figure 3, we can see what is the impact in the images for the SSIM metric. It is worth noticing that, for the top and bottom rows of the equirectangular image, the number of pixels that can be removed with minimal information loss is significant. It is possible to remove very large gaps of pixels in those regions and reconstruct them with minimal quality loss.

We colored both plots in Figure 2 and 3 to reveal how the metrics transition from high redundancy to low redundancy regions. To do so, we used the histogram of each metric's data (Figure 4a and 4b) to decide the range of the transition data. In the PSNR case we selected 65dB to be black color to differentiate the two regions, but in the SSIM case, there is no clear boundary and the transition is smoother.

The results from the WPSNR metric (Figure 5) shows a similar behavior as the PSNR with a similar histogram (Figure 7a). The S-SSIM, however, shows a completely
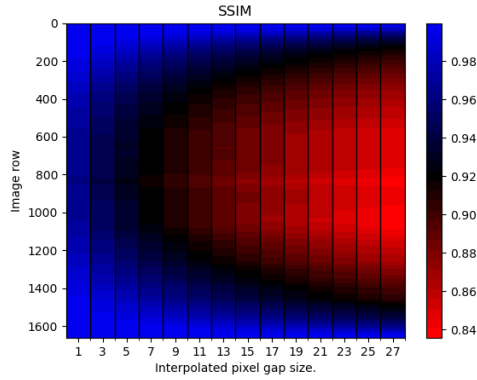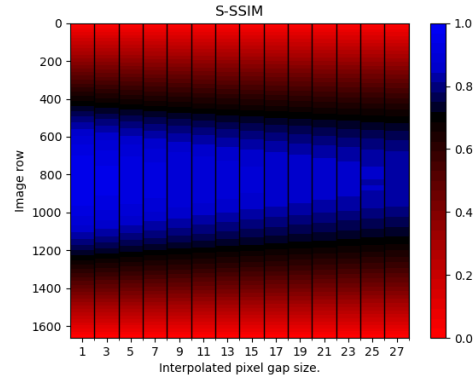
Figure 3: SSIM results.
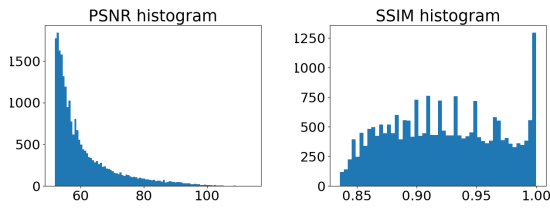

Figure 6: S-SSIM results.

(a)   (b)

Figure 4: (a) PSNR histogram and (b) SSIM histogram.

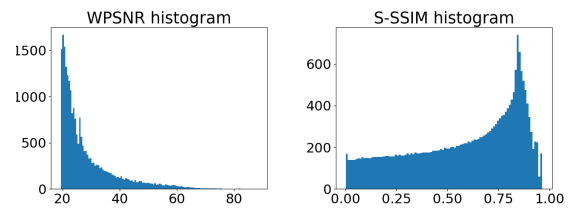(a)   (b)

Figure 7: (a) WPSNR histogram and (b) S-SSIM histogram.
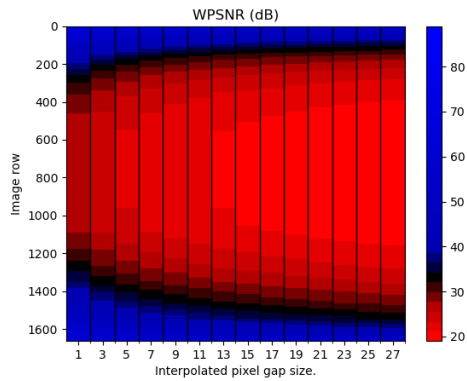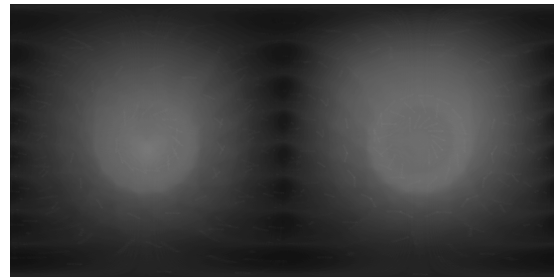

Figure 5: WPSNR results.

Figure 8: ROAD results, the values are scaled for better visualization.

different data from the standard SSIM. Highly redundant regions in SSIM vanish in S-SSIM due to the cosine factor of the weighting (Equation 1). However, it is possible to notice in Figure 6 a linear perceptual quality decrease towards the middle rows as long as the size of the interpolated pixel gap increases.

As expected from the equirectangular projection, these results show how the visual alterations of these type of images have a consine behavior of relevant pixels. However, these results do not show how the relevant pixels are structured across the image.

The pixel impact analysis produced interesting results on how the redundancy is scattered across equirectangular images. The computation of the localized SSIM metric shows, in Figure 9, the redundancy structure when projecting a 360° photo using the equirectangular projection. In this resulting image, dark pixels rep-

resent highly redundant regions and light pixel regions where the redundancy is minimum. The results reveal two circular regions centered in the longitudes $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ of the sphere that contain low redundancy and fade as we move further from their centers.

It is important to notice the existence of vertical dark lines; they represent the pixels used as reference for the interpolation and therefore they have no error when computing the metric values. The behaviour of these lines at the top and bottom rows of the image is interesting too. They get thicker as long as they get closer to the boundaries of the image due to the massive pixel stretching of the equirectangular projection in those regions.

The results from the ROAD metric (Figure 8) share a similar behavior as the SSIM. The difference of these two metrics is that the ROAD metric, as stated in Section 3.2.1, measures the gradients of a 3×3 region without any interpolation. This reveals the same circular
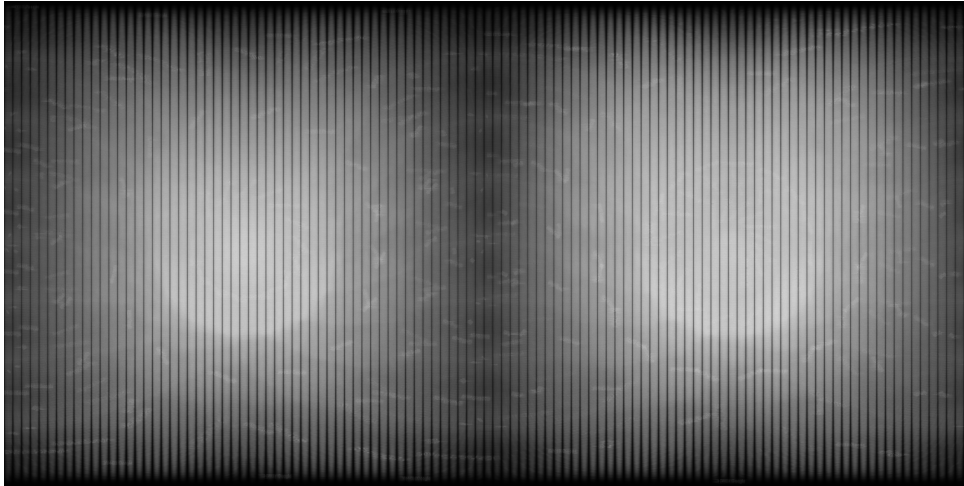
Figure 9: SSIM region results for 25 pixel gap size. Notice that the vertical black lines show the reference pixels from the interpolation where the error is non existent. Close to the top and bottom regions, the lines get thicker due to the highly redundant regions. It is important to mention that the images contain watermarks and are visible in the results.

low redundancy regions centered at longitudes $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. This means that those regions are the less distorted ones when we project the 3D sphere to the 2D plane using the equirectangular projection.

6 CONCLUSIONS

In this paper, we presented an analysis of pixel redundancy in equirectangular images. We analyzed the behavior of the visual quality assessment using standard metrics for regular 2D images and equirectangular images (PSNR and SSIM). In addition to that, we used the $360°$ image focused versions to complete the analysis and produce meaningful results. After removing several sets of pixels from rows and reconstructing them, we have shown how each metric behaves in a row-wise manner. The results showed consistency with previous statements of diverse authors and provided a visual demonstration of how the redundancy is increased at angles farther from the horizon line in equirectangular images. We revealed the redundancy structure of equirectangular images by using localized versions of SSIM and ROAD metrics, showing how sample values are stretched across pixels in images that use equirectangular projection. These results set a support for future research work in $360°$ image compression by providing how each portion of the equirectangular image contributes to the image quality. The knowledge of the redundancy structure in equirectangular images is useful to improve state of the art compression algorithms and improve the visual quality and immersion sensation in VR devices.

7 REFERENCES

[Aze19] Roberto Gerson Azevedo. Visual distortions in 360-degree videos. *IEEE Trans.* *Circuits Syst. Video Technol.*, PP(99), July 2019.

[Boy17] Jill Boyce. Spherical rotation orientation indication for HEVC and JEM coding of 360 degree video. In *Applications of Digital Image Processing XL*, volume 10396, page 103960I. International Society for Optics and Photonics, September 2017.

[Bud15] M Budagavi. 360 degrees video coding using region adaptive smoothing. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 750–754. ieeexplore.ieee.org, September 2015.

[Cha18] Shih-Hsiu Chang. Generating 360 outdoor panorama dataset with reliable sun position estimation. In *SIGGRAPH Asia 2018 Posters*, number Article 22 in SA '18, pages 1–2, New York, NY, USA, December 2018. Association for Computing Machinery.

[Che18] S Chen. Spherical structural similarity index for objective omnidirectional video quality assessment. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. ieeexplore.ieee.org, July 2018.

[DA17] A De Abreu. Look around you: Saliency maps for omnidirectional images in VR applications. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. ieeexplore.ieee.org, May 2017.

[DS16] F De Simone. Geometry-driven quantization for omnidirectional image coding.

In *2016 Picture Coding Symposium (PCS)*, pages 1–5. ieeexplore.ieee.org, December 2016.

[Dua18] H Duan. Perceptual quality assessment of omnidirectional images. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. ieeexplore.ieee.org, May 2018.

[Faced] Facebook. *Facebook 360 video*, 2021 (last accessed).

[Gar05] Roman Garnett. A universal noise removal algorithm with an impulse detector. *IEEE Trans. Image Process.*, 14(11):1747–1754, November 2005.

[Gooed] Google. *Tilt Brush by Google*, 2021 (last accessed).

[Gut18] Jesús Gutiérrez. Toolbox and dataset for the development of saliency and scanpath models for omnidirectional/360 still images. *Signal Processing: Image Communication*, 69:35–42, 2018.

[Hua18] Mingkai Huang. Modeling the perceptual quality of immersive images rendered on head mounted displays: Resolution and compression. *IEEE Trans. Image Process.*, August 2018.

[Jam19] M Jamali. Comparison of 3D 360-degree video compression performance using different projections. In *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, pages 1–6. ieeexplore.ieee.org, May 2019.

[Lee17] S-H Lee. Omnidirectional video coding using latitude adaptive down-sampling and pixel rearrangement. *Electron. Lett.*, 53(10):655–657, April 2017.

[McCed] John McCormac. *SceneNet, https://robotvault.bitbucket.io/scenenet-rgbd.html*, 2020 (last accessed).

[oITed] Visual Computing Lab Institute of Information Technology. *3D60 Dataset, https://vcl.iti.gr/360-dataset/*, 2020 (last accessed).

[oTed] Massachusetts Institute of Technology. *SUN 360 dataset, http://people.csail.mit.edu/jxiao/SUN360/*, 2020 (last accessed).

[Sit18] Vincent Sitzmann. Saliency in VR: How do people explore virtual environments? *IEEE Trans. Vis. Comput. Graph.*, 24(4):1633–1642, April 2018.

[Soned] Shuran Song. *SUNCG, https://sscnet.cs.princeton.edu/*, 2020 (last accessed).

[Sun17] W Sun. CVIQD: Subjective quality evaluation of compressed virtual reality images. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3450–3454. ieeexplore.ieee.org, September 2017.

[Sun18] W Sun. A Large-Scale compressed 360-degree spherical image database: From subjective quality evaluation to objective model comparison. In *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6, August 2018.

[Sze10] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[Wan04] Zhou Wang. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, April 2004.

[You16] R G Youvalari. Analysis of regional down-sampling methods for coding of omnidirectional video. In *2016 Picture Coding Symposium (PCS)*, pages 1–5. ieeexplore.ieee.org, December 2016.

[Yu15] M Yu. A framework to evaluate omnidirectional video coding schemes. In *2015 IEEE International Symposium on Mixed and Augmented Reality*, pages 31–36. ieeexplore.ieee.org, September 2015.