A PERCEPTUAL VISIBILITY METRIC FOR BANDING ARTIFACTS

Yilin Wang, Sang-Uok Kum, Chao Chen, Anil Kokaram

Google Inc., 1600 Amphitheatre Pkwy., Mountain View, CA, USA 94043
{yilin,kumsu,chaochen,anilkokaram}@google.com

ABSTRACT

Banding is a common video artifact caused by compressing low texture regions with coarse quantization. Relatively few previous attempts exist to address banding and none incorporate subjective testing for calibrating the measurement. In this paper, we propose a novel metric that incorporates both edge length and contrast across the edge to measure video banding. We further introduce both reference and non-reference metrics. Our results demonstrate that the new metrics have a very high correlation with subjective assessment and certainly outperforms PSNR, SSIM, and VQM.

Index Terms— Video Quality Measurement, Banding, Video Compression Artifacts, Visual Perception

1. INTRODUCTION

The automated measurement of video quality has been of interest to the video processing community for some time. With the increasing importance of OTT (over the top) streaming services [1], there has been a focus on the visual perception of artifacts specifically related to compression [2]. While the perception of detail loss, blockiness and to some extent mosquito noise artifacts have received a large amount of attention, there has been relatively little work on the perception of banding artifacts. This artifact is becoming increasingly important with the increasing fidelity of displays and improving content quality streamed to users. Banding artifacts (also called false contours) usually occur when quantizing the intensity/color of a low texture region (e.g. sky and water) with insufficient bits, and then the entire region seems to be formed by multiple bands with gradually changed intensity/color. An example of banding is shown in figure 1. The banding example was created as a result of the effect of transcoding the original uploaded clip in one of the YouTube ingest pipelines. In a subjective assessment exercise we measure a huge difference in MOS (Mean Opinion Scores) for the quality of the transcoded clip (50 units) with respect to the uploaded original. Yet traditional metrics like PSNR and SSIM [3] rate the quality of the transcode as 0.99 and 50dB respectively. Clearly therefore the traditional metrics do not align well with the subjective perception of this artifact.

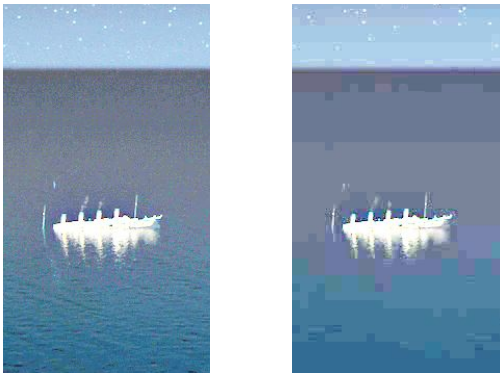

Fig. 1. *Banding artifacts caused by transcoding. Left:Frame extracted from a clip in a movie on YouTube. Right: Corresponding transcoded frame (served from YouTube). In this case PSNR, SSIM, and VQM are 50.16, 0.99, and 0.01 respectively. Our subjective test on the whole clip shows a huge MOS difference between original (85.33) and transcoded (34.61) video. These metrics therefore do not reflect the observed perception of banding artifacts.*

In this paper we develop a new algorithm for the detection of banding in both reference and non-reference cases. The essence of our algorithm is the delineation of homogenous segments (defined as "Unisegs") in the observed frame. The size of segments as well as the relative brightness between segments is used to assess the visibility of the artifact. Improving on the earlier work in [4, 5, 6] we present a more computationally efficient process. In addition, we calibrate the metric against subjective assessment and report on the mapping between the metric and observed MOS scores. We show our metric to be much better correlated with human perception than PSNR, SSIM or VQM[7]. In the next section we give some background on the topic of banding detection and then we go on to present our new detector and its assessment.

2. BACKGROUND

As far as we know there are just five previous published attempts to consider banding artifacts in some way [4, 5, 6, 8, 9]. While [8, 9] do not consider the measurement of band-

ing artifacts, the others [4, 5, 6] all employ the detection of uniform segments in their measurements. The challenge is to identify a boundary as indicative of banding (a "false contour") rather than a "true" region edge in the image, and such boundaries cannot be easily found by traditional edge detectors (e.g. Canny). Figure 2 shows the edge map for an image with visible banding artifacts, which contains so much edge clutter that it becomes very difficult to find a clean uniseg boundary. Thus traditional edge detectors cannot be directly used for detecting the uniseg boundary.

Baugh et al. [6] related the number of unisegs to the visibility of banding. Their observation was that when the size of most of the unisegs in an image was less than 10 pixels in area, then there was no perceivable banding in the image. However, the visibility of this artifact is related not just to the size of the uniseg but also to the relative contrast of the boundary pixels. Their work did not consider how a metric related to the number of unisegs was correlated to the perception of banding, rather they used this as an uncalibrated indicator of banding.

Lee et al. [4] segmented the image into smooth and non-smooth regions, and computed various directional features (e.g. contrast and Sobel masks) and non-directional features (e.g. Variance and Entropy) for each pixel in the non-smooth region to identify it as "banding boundaries" or "true edges". However, the experiment results showed that no single feature had a good correlation with all test cases.

Bhagavathy et al. [5] identified uniseg boundaries using the intensity distribution in a block centered at each pixel. For each block size (1 to 20 pixels), the likelihood of banding is estimated as the ratio of pixels that differ by 1 from the intensity at the current site. The block with the highest banding likelihood was used to determine whether there was significant banding. However, computing banding likelihoods for all pixels at all neighborhood scales is very expensive.

In our work we address the shortcomings of previous efforts by incorporating a subjective study into our assessment, reducing the complexity of the process and proposing both reference and non-reference metrics for this artifact.

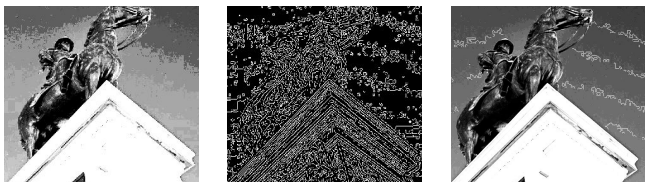

Fig. 2. *Cropped input frame (left) and corresponding uniseg boundaries detected by Canny (middle) and proposed method (right).*

3. BANDING FEATURES

We observe that the two essential features that affect the visibility of banding are i) the size or area of the bands and ii) the contrast across the banding contour. It turns out that the length of the contour is a good enough proxy to capture size information, and the contrast can be measured by the intensity coherence around the banding contour.

To find all banding contours, we first find all 4-connected unisegs in the frame. The segmentation is primitive in the sense that we look for absolutely no gradient in a uniseg i.e. connect pixels having exactly the same intensity. This directly surfaces contours likely to show banding and considerably reduces edge clutter. Small segments (having a size $\leq 0.2\%$ of the area of the image) are rejected from consideration. The boundaries of remaining unisegs are collected as possible banding contours. Since one uniseg may have multiple neighboring unisegs, its boundary will be further split into multiple edges (called Banding Edges), where each banding edge correspond to one neighboring uniseg. Figure 2 shows how this process improves upon the use of traditional edge detection for this purpose.

Define $I(p)$ as the intensity at a pixel site p and $N(p)$ as the set of pixels within some neighborhood of p. Let B denote a mask in which all pixels belonging to the boundary of the ith uniseg are labeled as i, and 0 otherwise. A site p between the ith and jth unisegs is labeled as a banding edge pixel, $E_{i,j}(p) = 1$, when:

$$E_{i,j}(p) = \begin{cases} 1 & \text{if } B(p) = i \text{ and } Q^p_{i,j} \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Where $Q^p_{i,j} = N(p) \cap \{q | B(q) = j,\ 0 < I(p) - I(q) < T\}$, and T is a threshold ($= 5$ here) that rejects edges having a very high contrast and hence interpreted as true edges and not banding edges. We use a block of 11×11 pixels as our neighborhood. The size feature for banding edge $E_{i,j}$ between two unisegs is its cardinality $|E_{i,j}|$. However, this feature is not good enough to measure banding visibility. Figure 3, shows bands on both frames, and the lengths of banding edges on both frames are similar, but obviously the banding degrees of those two frames are different.

To address this, we separate pixels outside a uniseg but within the neighborhood of a banding edge into two groups: one group has the same intensity as the banding edge, and the other group has different intensity. The ratio of the two group sizes is defined as a contrast feature, called Edge Coherence. The intuition is shown in Figure 3. The right image shows a banding edge in the green/blue boundary which contains intensities within the green uniseg that are the same as in the blue uniseg. The banding edge on the right is "dithered" and so is less visible.

Suppose M is the uniseg map where pixels belonging to the ith uniseg is marked as i. The edge coherence for banding

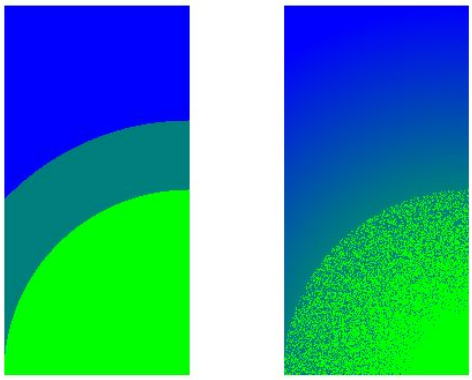

Fig. 3. *Edge Coherence Comparison. The green uniseg in both frames have visible boundaries, but their strength (banding coherence) are different.*

edge $E_{i,j}$, denoted by $C_{i,j}$ is defined as:

$$C_{i,j} = 1 - \min\left(1, \frac{|\{q|I(q) = I(p)\} \cap O^p_{i,j}|}{|\{q|I(q) \neq I(p)\} \cap O^p_{i,j}|}\right)$$

where $p \in E_{i,j}$ and $O^p_{i,j} = N(p) \cap \{q|M(q) \neq i\}$.

So far we defined two features: banding edge length and edge coherence. In next section, we will use them to design both non-reference and reference banding metrics.

4. BANDING METRICS AND THE PIPELINE

The pipeline consists of four components: uniseg generation, banding edge detection, reference banding edge matching, and banding score evaluation, as shown in Figure 4. In our system the output YouTube transcode is used to generate unisegs. That map typically contains large regions in which there are unisegs which are rejected as being too small. The uniseg boundaries are further split into banding edges. In the example shown in Figure 5 we can see there are no banding edges between the sky and the building, because there are no unisegs on the building. Each banding edge fragment is then assigned a length and edge coherence value as in the previous section. We also measure the contrast $r_{i,j}$ across each banding fragment $E_{i,j}$ as the difference in intensity between the unisegs i and j.

To generate a reference banding score Q^b_r we use banding edge detection in the transcoded frames to delineate regions of interest in the uploaded (original) frames. The corresponding regions are compared and we reject banding edges in the transcode which show edge coherence similar to the coherence of edges in the original. The remaining edges in the transcode are then used to generate Q^b_r.

Given banding edges detected in the transcode, we therefore project the region around them (including the relevant

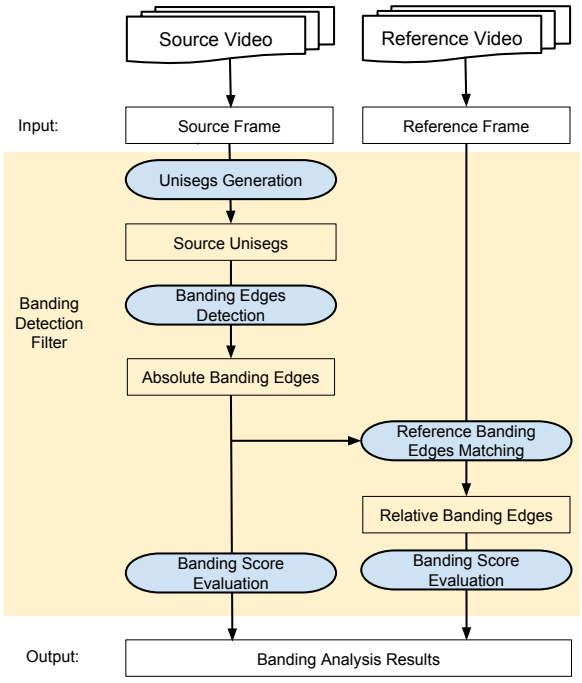

Fig. 4. *Banding Detection Pipeline.*

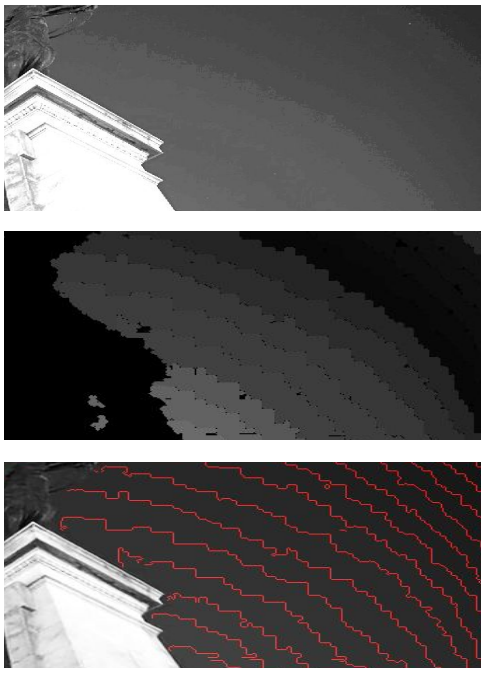

Fig. 5. *Input source frame (top), uniseg map (middle), and banding edge map (bottom).*

neighbourhood) into the source. This also allows us to account for different image sizes in transcode and source due for instance to downsampling operations required for creation of DASH [10] representations in our transcoding pipeline. Uniseg extraction then proceeds in the source, and we use banding edges for large unisegs in the source as the reference banding edges. When a reference banding edge is found, the difference of the source and transcode edge coherence is measured. If the difference is less than the visible coherence difference ($= 0.2$ here) then we reject the banding edge in the transcode since it is visibly similar to that in the source. We keep all edges in the transcode which do not exist in the source since these have been introduced by transcoding and hence will affect Q_r^b.

The final reference score is then calculated over the remaining edge fragments in the transcode. Fragments are further rejected if $C_{i,j} < 0.95$ our threshold for fragment visibility. The score then is $Q_r^b = \sum(1/D)|E_{i,j}| \times r_{i,j}$ over the whole image. We use $1/D$ as a normalizing factor for the image size where D is the length of the diagonal. For the video clips we use in our experiment the score is the average Q_r^b over the clip.

Our non-reference score Q_n^b therefore is simply the value above but calculated over all banding edges in an image. Hence we once more detect banding edge fragments but reject these only on the basis of visibility in that frame ($C_{i,j} < 0.95$) and not on the basis of comparison to the source.

5. SUBJECTIVE EVALUATION

Subjective assessment was performed by allowing subjects to rate two video clips shown side by side exhibiting different levels of banding. To generate these levels we transcoded 7 clips of 720p 30fps material with different levels of quantization (QP in ffmpeg) using VP9 [11]. Three levels of banding were selected from each set of transcodes. Subjects then assessed 7×6 sets of side by side comparisons. The clips were rated according to the usual BT500 standard yielding MOS for each banding level example. We used a 55 inch Samsung 4K Ultra HD TV for assessment with 25 subjects, and the test took about 1 hour for each participant. Figure 6 shows the MOS for all our clips compared to the scoring metric. The plots were generated from over 1000 measurements made through our test. As we can see the correlation is very acceptable: for Q_n^b, banding score greater than 5 corresponds to MOS lower than 40, which in some sense means the videos have unacceptable banding artifacts. For the reference case Q_r^b, a score greater than 5 corresponds to DMOS (Difference MOS) lower than -20, which means the transcoding process introduced unacceptable banding artifacts. The fit between MOS and our metrics is based on an exponential model: $MOS = c_0 + c_1 \cdot EXP(c_2 \cdot Q^b)$, where $(c_0,\ c_1,\ c_2)$ is $(14.485,\ 58.306,\ -0.140)$ for the non-reference case and $(-50.690,\ 48.630,\ -0.206)$ for the reference case (shown as the dotted curves in Figure 6). The MOS score generated in this way has a correlation of 0.849 (non-reference) and 0.883 (reference) with the observed MOS score for reference and non-reference metrics.

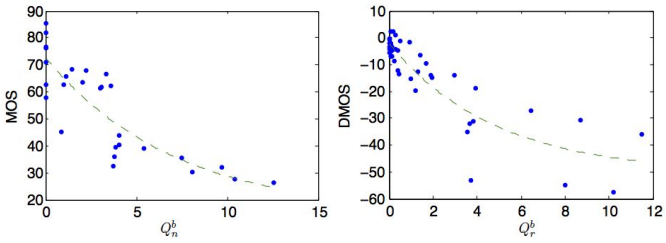

Fig. 6. *Correlation between proposed banding scores and subjective scores for non-reference (left) and reference (right) cases. The dotted curves are generated by exponential fitting.*

	PSNR	SSIM	VQM	Baugh[6]	Proposed
MOS	0.339	0.109	0.205	0.750	0.821
DMOS	0.512	0.353	0.353	0.759	0.845

Table 1. *Linear Correlation between Subjective Scores and Objective Metrics*

The result was also compared against Baugh's banding metric proposed in [6] as well as three general objective metrics: PSNR, SSIM, and VQM. Non-reference scores for general metrics were computed by using the original video as reference video. Reference scores were the difference between two corresponding non-reference scores. The linear correlation coefficients are shown in Table 1, we can see the proposed banding metric has better correlation with the subjective score than Baugh's metric and general metrics PSNR, SSIM, and VQM. The cross validation errors for both non-reference and reference banding metrics are less than 10 (in scale of 100), which means that most predicted scores and their corresponding true scores are in the same banding level.

6. CONCLUSION

In this paper, we present new algorithms for the visual perception of banding artifacts. The key features used in the new metrics are the contrast across banding edges and the length of these edges. An important aspect of our process is the use of primitive segmentation to analyze only boundaries which are likely to represent banding. Our results show correlation of better than 0.8 with subjective assessments which compares very favorably with existing general metrics. The banding detector proposed in this paper treats each frame independently, which could be further improved by incorporating temporal information. This is an important feature and we will address this in future work.

7. REFERENCES

[1] Anil Kokaram, Regis Crinon, and Nicolas Catania, "Ott (over-the-top)," *SMPTE Motion Imaging Journal*, vol. 124, pp. 65–68, 2015.

[2] Muhammad Shahid, Andreas Rossholm, Benny Lövström, and Hans-Jürgen Zepernick, "No-reference image and video quality assessment: a classification and review of recent approaches," *EURASIP J. Image and Video Processing*, vol. 2014, pp. 40, 2014.

[3] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, 2004.

[4] Ji Won Lee, Bo Ra Lim, Rae-Hong Park, Jae-Seung Kim, and Wonseok Ahn, "Two-stage false contour detection using directional contrast features and its application to adaptive false contour reduction," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 1, 2006.

[5] Sitaram Bhagavathy, Joan Llach, and Jie fu Zhai, "Multi-scale probabilistic dithering for suppressing banding artifacts in digital images," in *ICIP*, 2007.

[6] Gary Baugh, Anil Kokaram, and Francois Pitie, "Advanced video debanding," in *European Conference on Visual Media Production (CVMP)*, 2014.

[7] Margaret H. Pinson and Stephen Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, 2004.

[8] Scott J. Daly and Xiaofan Feng, "Decontouring: prevention and removal of false contour artifacts," in *Proc. SPIE 5292, Human Vision and Electronic Imaging IX*, 2004.

[9] Xin Jin, S. Goto, and K. N. Ngan, "Composite model-based dc dithering for suppressing contour artifacts in decompressed video," *IEEE Transactions on Image Processing*, vol. 20, no. 8, 2011.

[10] Thomas Stockhammer, "Dynamic adaptive streaming over http – standards and design principles," *Proceedings of the Second Annual ACM Conference on Multimedia Systems*, pp. 133–144, 2011.

[11] Debargha Mukherjee, Jim Bankoski, Jingning Grange, Adrian abd Han, John Koleszar, Paul Wilkins, Yaowu Xu, and Ronald Bultje, "The latest open-source video codec vp9 - an overview and preliminary results," *Picture Coding Symposium (PCS)*, pp. 390–393, 2013.