

BBAND INDEX: A NO-REFERENCE BANDING ARTIFACT PREDICTOR

Zhengzhong Tu*, Jessie Lin[†], Yilin Wang[†], Balu Adsumilli[†], and Alan C. Bovik*

*Laboratory for Image and Video Engineering (LIVE), The University of Texas at Austin

[†]YouTube Media Algorithms Team, Google Inc.

ABSTRACT

Banding artifact, or false contouring, is a common video compression impairment that tends to appear on large flat regions in encoded videos. These staircase-shaped color bands can be very noticeable in high-definition videos. Here we study this artifact, and propose a new distortion-specific no-reference video quality model for predicting banding artifacts, called the **Blind BANDING** Detector (BBAND index). BBAND is inspired by human visual models. The proposed detector can generate a pixel-wise banding visibility map and output a banding severity score at both the frame and video levels. Experimental results show that our proposed method outperforms state-of-the-art banding detection algorithms and delivers better consistency with subjective evaluations.

Index Terms— Video quality predictor, compression artifact, banding artifact, false contour, human visual model

1. INTRODUCTION

Banding/false contour remains one of the dominant artifacts that plague the quality of high-definition (HD) videos, especially when viewed on high-resolution or Retina displays. Yet, while significant research effort has been devoted to analyzing various specific compression related artifacts [1], such as noise [2], blockiness [3], ringing [4], and blur [5], less attention has been paid to analyzing banding/false contours. Given the rapidly growing demand for HD/Ultra-HD videos, the need to assess and mitigate banding artifacts is receiving increased attention in both academia and industry.

Banding appears in large, smooth regions with small gradients and presents as discrete, often staircased bands of brightness or color as a result of quantization in video compression. All popular video encoders, including H.264/AVC [6], VP9 [7], and H.265/HEVC [8] can introduce these artifacts at lower or medium bitrate when coding contents containing smooth areas. Fig. 1 shows an example of banding artifacts exacerbated by transcoding. Traditional quality prediction algorithms such as PSNR, SSIM [9], and VMAF [10], however, do not align well with human perception of banding [11]. The development of a highly reliable banding detector for both original user-generated content (UGC) and

*Work done during an internship at Google.

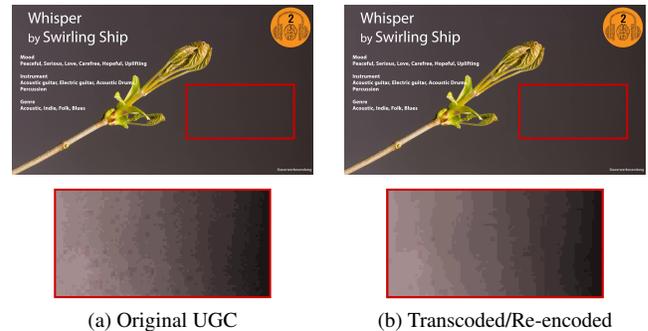


Fig. 1: Banding artifacts exacerbated by transcoding/re-encoding. (a) shows a frame sampled from an original UGC video with less noticeable “noisy” banding edges, while VP9-encoding exhibits more visible “clean” banding edges, as shown in (b). The lower figures show contrast-enhanced banding regions for better visualization.

transcoded/re-encoded videos, would, therefore, greatly assist streaming platforms in developing measures to avoid banding artifacts in streaming videos.

Related Work. There exists some prior study relating to banding/false contour detection. Some methods [12–14] take advantage of local features such as the gradient, contrast or entropy to measure potential banding edge statistics. However, methods like these generally do not perform very well when applied to assess the severity of banding edges in video content. Another approach to banding detection is based on pixel segmentation [11, 15, 16], where a bottom-up connected component analysis is used to first detect uniform segments, usually followed by a process of banding edge separation. These methods are often sensitive to edge noise, though. We do not include block-based processing, as in [17, 18], since it is hard to classify blocks where banding and textures coexist. If post-filtering is applied to these blocks, textures near the banding may become over-smoothed.

Our objective is to design an adaptive blind processor which can detect or enhance both “noisy” banding artifacts that arise in original UGC videos, as well as “clean” banding edges in transcoded videos. In this regard, it could be utilized as a basis for the development of pre-processing and

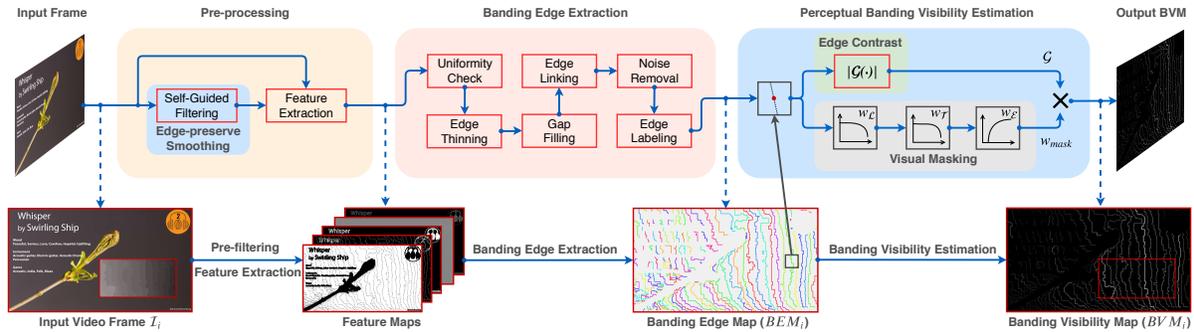


Fig. 2: Schematic overview of the first portion (Section 2.1-2.3) of the proposed BBAND model. The first row shows the processing flow, while the second row depicts exemplar responses of each processing block.

post-processing *debanding* algorithms. More recent banding detectors like the False Contour Detection and Removal (FCDR) [14] and Wang’s method [11] are not designed for this practical purpose, and hence it is essential to devote more research to developing other adaptive banding predictors applicable to pre- or post-debanding implementations.

In this paper we propose a new, “completely blind” [19] banding model, dubbed the **Blind BANDING Artifact Detector** (BBAND index), by leveraging edge detection and a human visual model. The proposed method operates on individual frames to obtain a pixel-wise banding visibility map. It can also produce no-reference perceptual quality predictions of videos with banding artifacts. Details of our proposed banding detector are given in Section 2, while evaluation results are given in Section 3. Finally, Section 4 concludes the paper.

2. PROPOSED BANDING DETECTOR

A block diagram of the first portion of the proposed model, which generates a pixel-wise banding visibility map (BVM), is illustrated in Fig. 2. Based on our observation that banding artifact appears as weak edges with small gradient (whether “clean” or “noisy”), we build our banding detector (BBAND), by exploiting existing edge detection techniques as well as certain visual properties. A spatio-temporal visual importance pooling is then applied to the BVM, as shown in Fig. 3, yielding “completely blind” banding scores for both individual frames and the entire video.

2.1. Pre-processing

We have observed that re-encoding videos at bitrates optimized for streaming often exacerbates banding in videos that already exhibit slight banding artifacts that may be barely visible, as shown in Fig. 1[§]. We thereby deployed self-guided filtering [20], which is an effective edge-preserving smoothing process, to enhance banding edges. We deemed the guided

[§]The exemplary frames used in this paper are from Music2Brain (YouTube channel). Website: <https://bigweb24.de>. Used with permission.

to be a better choice than the bilateral filter [21], since it better preserves gradient profile, which is a vital local feature used in our proposed framework. Image gradients are then calculated by applying a Sobel operator after pre-smoothing, yielding a gradient feature map.

2.2. Banding Edge Extraction

Inspired by the efficacy of using the Canny edge detector [22] to improve ringing region detection [23], we performed a similar procedure to extract banding edges. After pre-filtering, the pixels are classified into three classes depending on their Sobel gradient profiles: pixels having Sobel gradient magnitudes less than T_1 are labeled as flat pixels; pixels with gradient magnitudes exceeding T_2 are marked as textures. The remaining pixels are regarded as candidate banding pixels (CBP), on which the following steps are implemented to create a banding edge map (BEM). (We used $(T_1, T_2) = (2, 12)$).

1. *Uniformity Check*: Only the CBPs whose neighbors are either flat pixels or CBPs are retained for further processing.
2. *Edge Thinning*: Non-maxima suppression [22] is applied to each remaining CBP along its Sobel gradient orientation to better localize the potential bands.
3. *Gap Filling*: If two candidate pixels are disjoint, but able to be overlapped by a binary circular blob, the gap between the two points is filled by a proper banding edge.
4. *Edge Linking*: All connected CBPs are linked together in lists of sequential edge points. Each edge is either a curved line or a loop.
5. *Noise Removal*: Linked edges shorter than a certain threshold are discarded as visually insignificant.
6. *Edge Labeling*: The resulting connected banding edges are labeled separately, defining the ultimate BEM.

The colored edge map in Fig. 2 shows a BEM extracted from an input frame. The banding edges are well localized.

2.3. Banding Visibility Estimation

Staircase-like banding artifacts appear similar to Mach Bands (Chevreul illusion), where perceived edge contrast is ex-

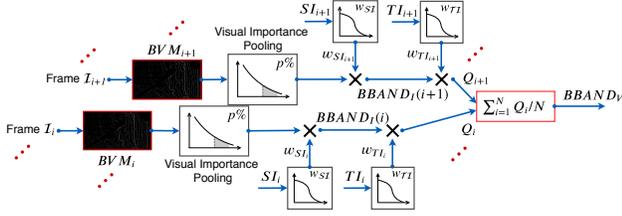


Fig. 3: Flowchart of the second portion (Section 2.4) of the proposed BBAND model, which produces banding scores on both frames and whole videos.

aggregated by edge enhancement by the early visual system [24]. Explanations of the illusion usually involve the center-surround excitatory-inhibitory pooling responses of retinal receptive fields [25]. Inspired by the psychovisual findings in [26], we developed a local banding visibility estimator based on edge contrast and perceptual masking effects. The estimator processes the BEM and yields an element-wise banding visibility map (BVM).

2.3.1. Basic Edge Feature

Banding artifact presents as visible edges. As described earlier, we use the Sobel gradient magnitude as an edge visibility feature. Since edge visibility is also affected by content, we also model visual masking as it may affect the subjective perception of banding.

2.3.2. Visual Masking

Visual masking is a phenomenon whereby the visibility of a visual stimulus (target) is reduced by the presence of another stimulus, called a mask. Well-known masking effects include luminance and texture masking [23, 27]. Here we deploy a simple but effective quantitative model of the effect of masking on banding edge visibility.

Local Statistics: At each detected banding pixel in the BEM, compute local Gaussian-weighted mean and standard deviation (“sigma field”) on the original un-preprocessed frame:

$$\mu(i, j) = \sum_{k=-K}^K \sum_{\ell=-L}^L w_{k, \ell} \mathcal{I}(i-k, j-\ell) \quad (1)$$

$$\sigma(i, j) = \sqrt{\sum_{k=-K}^K \sum_{\ell=-L}^L w_{k, \ell} [\mathcal{I}(i-k, j-\ell) - \mu(i, j)]^2}, \quad (2)$$

where (i, j) are spatial indices at detected pixels in the BEM with corresponding original pixel intensity $\mathcal{I}(i, j)$, and $\mathbf{w} = \{w_{k, \ell} | k = -K, \dots, K, \ell = -L, \dots, L\}$ is a 2D isotropic Gaussian weighting function. We use the $\mu(\cdot)$ and $\sigma(\cdot)$ feature maps to estimate the local background luminance and complexity. The window size in our experiments was set as 9×9 .

Luminance Masking: We define a luminance visibility transfer function (VTF_ℓ) to express luminance masking as a function of the local background intensity. We have observed that banding artifacts remain visible even in very dark areas, so we only model the masking at very bright pixels. A final luminance masking weight is computed at each pixel as

$$w_\ell(i, j) = \begin{cases} 1 & \mu(i, j) \leq \mu_0 \\ 1 - \alpha(\mu(i, j) - \mu_0)^\beta & \mu(i, j) > \mu_0, \end{cases} \quad (3)$$

where $\mu(i, j)$ is calculated using (1). (α, β) is a pair of constants chosen to adjust the shape of the transfer function. We used $(\alpha, \beta, \mu_0) = (1.6 \times 10^{-5}, 2, 81)$ in our implementations.

Texture Masking: We also define a texture visibility transfer function (VTF_t) to capture the effects of texture masking. The VTF_t is defined to be inversely proportional to local image activity [23] when an activity measure (mean “sigma field”) rises above threshold λ_0 . The overall weighting function is formulated as

$$w_t(i, j) = \begin{cases} 1 & \lambda(i, j) \leq \lambda_0 \\ 1/[1 + (\lambda(i, j) - \lambda_0)^\gamma] & \lambda(i, j) > \lambda_0, \end{cases} \quad (4)$$

and

$$\lambda(i, j) = \frac{1}{(2K+1)(2L+1)} \sum_{k=-K}^K \sum_{\ell=-L}^L \sigma(i-k, j-\ell), \quad (5)$$

where $\sigma(i, j)$ is given by Eq. (2), and γ is a parameter that is used to tune the nonlinearity of VTF_t . The values of $(\gamma, \lambda_0) = (5, 0.32)$ were adopted after careful inspection.

Cardinality Masking: The authors of [11] have shown that edge length is another useful banding visibility feature in a subjective study. We accordingly define the following transfer function which weights banding visibility by edge cardinality:

$$w_c(i, j) = \begin{cases} 0 & |E(i, j)| \leq c_0 \\ (|E(i, j)|/\sqrt{MN})^\eta & |E(i, j)| > c_0, \end{cases} \quad (6)$$

where $E(i, j) = \{E \in \text{BEM} | (i, j) \in E\}$ is the set of banding edges passing through location (i, j) , and c_0 is a threshold on minimal noticeable edge length, above which banding edge visibility is positively correlated to normalized edge length. M and N denote the image height and width, respectively. We used parameters $(c_0, \eta) = (16, 0.5)$ in our experiments.

2.3.3. Visibility Integration

The overall visibility of an artifact depends on the visual response to it modulated by a concurrency of masking effects. Here we use a simple but effective product model of feature integration at each computed banding pixel to obtain the banding visibility map:

$$\text{BVM}(i, j) = w_\ell(i, j) \cdot w_t(i, j) \cdot w_c(i, j) \cdot |\mathcal{G}(i, j)|, \quad (7)$$

where $w(\cdot)$ ’s are the responsive weighting parameters that scale the measured edge strength (Sobel gradient magnitude) $|\mathcal{G}(i, j)|$ at location (i, j) .

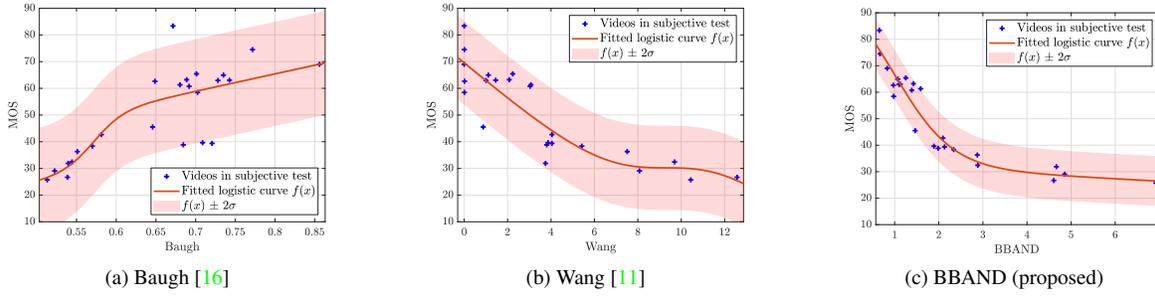


Fig. 4: Scatter plots and regression curves of (a) Baugh [16], (b) Wang [11], (c) BBAND, versus MOS on banding dataset [11].

Table 1: Performance comparison of blind banding models.

Metric	SRCC	KRCC	PLCC	RMSE
Baugh [16]	0.7739	0.6304	0.8037	9.7671
Wang [11]	0.8689	0.6788	0.8770	7.8863
BBAND	0.9330	0.8116	0.9578	4.7173

2.4. Making a Banding Metric

Previous authors [27–30] have studied the benefits of integrating visual importance pooling into objective quality model, generally aligning with the idea that the overall perceived quality of a video is dominated by those regions having the poorest quality. In our model, we apply the worst $p\%$ percentile pooling to obtain an average banding score from the extracted BVM, where $p = 80$ is employed in experiments.

Banding usually occurs in non-salient regions (e.g., background) while salient objects catch more of the viewer’s attention. We thereby use the well-known spatial information (SI) and temporal information (TI) to indicate possible spatial and temporal distractors against banding visibility. SI is computed as the standard deviation of the pixel-wise gradient magnitude, while TI as the standard deviation of the absolute frame differences on each frame [31]. These are then mapped by an exponential transfer function to obtain weights:

$$w_i(x) = \exp(-a_i x^{b_i}), \quad i \in \{\text{SI}, \text{TI}\}. \quad (8)$$

Finally, we construct the frame-level BBAND index by applying visual percentile pooling and SI weights to BVM:

$$Q_{\text{BBAND}_{\mathcal{I}}}(\mathcal{I}) = w_{\text{SI}}(\text{SI}) \cdot \frac{1}{|\mathcal{K}_{p\%}|} \sum_{(i,j) \in \mathcal{K}_{p\%}} \text{BVM}_{\mathcal{I}}(i, j), \quad (9)$$

where $\mathcal{K}_{p\%}$ is the index set of the largest p^{th} percentile non-zero pixel-wise visibility values contained in the BVM of frame \mathcal{I} . We also obtain the video-level BBAND metric by averaging all frame-level banding scores, weighted by per-frame TI, respectively:

$$Q_{\text{BBAND}_{\mathcal{V}}}(\mathcal{V}) = \frac{1}{N} \sum_{n=1}^N w_{\text{TI}}(\text{TI}_n) \cdot Q_{\text{BBAND}_{\mathcal{I}}}(\mathcal{I}_n). \quad (10)$$

Fig. 3 shows the entire workflow of the BBAND indices.

3. SUBJECTIVE EVALUATION

Other implemented parameters in our proposed BBAND model are $(a_{\text{SI}}, b_{\text{SI}}, a_{\text{TI}}, b_{\text{TI}}) = (10^{-6}, 3, 2.5 \times 10^{-3}, 2)$, respectively, after empirical calibration, and we’ve found these selected parameters generally perform well in most cases. We evaluated the BBAND model against two recent banding metrics, Wang [11] and Baugh [16], on the only existing banding dataset, created by Wang et al. [11]. It consists of six clips of 720p@30fps videos with different levels of quantization using VP9. The Spearman rank-order correlation coefficient (SRCC) and Kendall rank-order correlation coefficient (KRCC) between predicted scores and mean opinion scores (MOS) of subjects are directly reported for the evaluated methods. We also calculated the Pearson linear correlation coefficient (PLCC) and the corresponding root mean squared error (RMSE) after fitting a logistic function between MOS and predicted values [32]. Table 1 summarizes the experimental results, and Fig. 4 plots the fitted logistic curves of MOS versus the evaluated banding models. These results have shown that the proposed BBAND metric yields highly promising performance regarding subjective consistency.

4. CONCLUSION AND FUTURE WORK

We have presented a new no-reference video quality model called the BBAND for assessing perceived banding artifacts in high-quality or high-definition videos. The algorithm involves robust detection of banding edges, a perception-inspired estimator of banding visibility, and a model of spatial-temporal visual importance pooling. Subjective evaluation shows that our proposed method correlates favorably with human perception as compared to several existing banding metrics. As a “completely blind” (opinion-unaware) distortion-specific quality indicator, BBAND can be incorporated with other video quality measures as a tool to optimize user-generated video processing pipelines for media streaming platforms. Future work will include further improvements of BBAND by integrating with more temporal cues, and its applications to address such banding artifacts via debanding pre-processing or post-filtering.

5. REFERENCES

- [1] B. L. M. Shahid, A. Rossholm and H.-J. Zepernick, “No-reference image and video quality assessment: a classification and review of recent approaches,” *EURASIP J. Image Video Process.*, vol. 2014, no. 1, pp. 40–40, 2014.
- [2] A. Norkin and N. Birkbeck, “Film grain synthesis for av1 video codec,” in *2018 Data Compress. Conf.*, 2018, pp. 3–12.
- [3] Z. Wang, A. C. Bovik, and B. L. Evan, “Blind measurement of blocking artifacts in images,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 3, 2000, pp. 981–984.
- [4] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, “Perceptual blur and ringing metrics: application to jpeg2000,” *Signal Process. Image Commun.*, vol. 19, no. 2, pp. 163–172, 2004.
- [5] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, “A no-reference perceptual blur metric,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 3, 2002, pp. III–III.
- [6] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the h. 264/avc video coding standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, 2003.
- [7] D. Mukherjee, J. Bankoski, A. Grange, J. Han, J. Koleszar, P. Wilkins, Y. Xu, and R. Bultje, “The latest open-source video codec vp9-an overview and preliminary results,” in *Proc. Picture Coding Symp. (PCS)*, 2013, pp. 390–393.
- [8] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (hevc) standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [9] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [10] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, “Toward a practical perceptual video quality metric,” *The Netflix Tech Blog*, vol. 6, 2016.
- [11] Y. Wang, S. Kum, C. Chen, and A. Kokaram, “A perceptual visibility metric for banding artifacts,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2067–2071.
- [12] S. J. Daly and X. Feng, “Decontouring: Prevention and removal of false contour artifacts,” in *Proc. SPIE, Human Vision and Electron. Imag. IX*, vol. 5292, 2004, pp. 130–149.
- [13] J. W. Lee, B. R. Lim, R.-H. Park, J.-S. Kim, and W. Ahn, “Two-stage false contour detection using directional contrast and its application to adaptive false contour reduction,” *IEEE Trans. Consum. Electron.*, vol. 52, no. 1, pp. 179–188, 2006.
- [14] Q. Huang, H. Y. Kim, W.-J. Tsai, S. Y. Jeong, J. S. Choi, and C.-C. J. Kuo, “Understanding and removal of false contour in hevc compressed images,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 2, pp. 378–391, 2016.
- [15] S. Bhagavathy, J. Llach, and J. Zhai, “Multiscale probabilistic dithering for suppressing contour artifacts in digital images,” *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1936–1945, 2009.
- [16] G. Baugh, A. Kokaram, and F. Pitié, “Advanced video debanding,” in *ACM Proc. 11th Eur. Conf. Visual Media Prod.*, 2014, p. 7.
- [17] X. Jin, S. Goto, and K. N. Ngan, “Composite model-based dc dithering for suppressing contour artifacts in decompressed video,” *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2110–2121, 2011.
- [18] Y. Wang, C. Abhayaratne, R. Weerakkody, and M. Mrak, “Multi-scale dithering for contouring artefacts removal in compressed uhd video sequences,” in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, 2014, pp. 1014–1018.
- [19] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a completely blind image quality analyzer,” *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2012.
- [20] K. He, J. Sun, and X. Tang, “Guided image filtering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2012.
- [21] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, vol. 98, no. 1, 1998, p. 2.
- [22] J. Canny, “A computational approach to edge detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 6, pp. 679–698, 1986.
- [23] H. Liu, N. Klomp, and I. Heynderickx, “A perceptually relevant approach to ringing region detection,” *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1414–1426, 2010.
- [24] “Mach bands — Wikipedia, the free encyclopedia,” [Accessed 5-October-2019]. [Online]. Available: https://en.wikipedia.org/wiki/Mach_bands
- [25] F. Ratliff, *Mach bands: quantitative studies on neural networks*. Holden-Day, San Francisco London Amsterdam, 1965, vol. 2.
- [26] J. Ross, M. C. Morrone, and D. C. Burr, “The conditions under which mach bands are visible,” *Vision Research*, vol. 29, no. 6, pp. 699–715, 1989.
- [27] C. Chen, M. Izadi, and A. Kokaram, “A perceptual quality metric for videos distorted by spatially correlated noise,” in *ACM Multimedia Conf.*, 2016, pp. 1277–1285.
- [28] D. Ghadiyaram, C. Chen, S. Inguva, and A. Kokaram, “A no-reference video quality predictor for compression and scaling artifacts,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017, pp. 3445–3449.
- [29] A. K. Moorthy and A. C. Bovik, “Visual importance pooling for image quality assessment,” *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 193–201, 2009.
- [30] J. Park, K. Seshadrinathan, S. Lee, and A. C. Bovik, “Video quality pooling adaptive to perceptual distortion severity,” *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 610–620, 2012.
- [31] P. ITU-T RECOMMENDATION, “Subjective video quality assessment methods for multimedia applications,” *Int. Telecom. Union*, 1999.
- [32] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, 2006.