Accepted Manuscript

Video quality assessment accounting for temporal visual masking of local flicker

Lark Kwon Choi, Alan Conrad Bovik

PII: S0923-5965(18)30637-4
DOI: https://doi.org/10.1016/j.image.2018.06.009
Reference: IMAGE 15405

To appear in: *Signal Processing: Image Communication*

Received date : 10 July 2017
Revised date : 4 February 2018
Accepted date : 20 June 2018

Please cite this article as: L.K. Choi, A.C. Bovik, Video quality assessment accounting for temporal visual masking of local flicker, *Signal Processing: Image Communication* (2018), https://doi.org/10.1016/j.image.2018.06.009

Video Quality Assessment Accounting for Temporal Visual Masking of Local Flicker

Lark Kwon Choi *, Alan Conrad Bovik

Laboratory for Image and Video Engineering (LIVE), Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78701, USA

Abstract— An important element of the design of video quality assessment (VQA) models that remains poorly understood is the effect of temporal visual masking on the visibility of temporal distortions. The visibility of temporal distortions like local flicker can be strongly reduced by motion. Based on a recently discovered visual change silencing illusion, we have developed a full reference VQA model that accounts for temporal visual masking of local flicker. The proposed model, called Flicker Sensitive – MOtion- based Video Integrity Evaluation (FS-MOVIE), augments the well-known MOVIE Index by combining motion tuned video integrity features with a new perceptual flicker visibility/masking index. FS-MOVIE captures the separated spectral signatures caused by local flicker distortions, by using a model of the responses of neurons in primary visual cortex to video flicker, an energy model of motion perception, and a divisive normalization stage. FS-MOVIE predicts the perceptual suppression of local flicker by the presence of motion and evaluates local flicker as it affects video quality. Experimental results show that FS-MOVIE significantly improves VQA performance against its predecessor and is highly competitive with top performing VQA algorithms when tested on the LIVE, IVP, EPFL, and VQEGHD5 VQA databases.

Keywords—Video quality assessment, temporal visual masking, motion silencing, flicker visibility, human visual system.

1. Introduction

Digital videos have become pervasive in our daily life. Video streaming services such as Netflix and YouTube, video sharing in social media, and video calling using Skype have become commonplace. As mobile devices have become “smarter,” video consumption is exponentially increasing [1]. Given the dramatic growth in purveyed video content and heightened user expectations of higher-quality videos, it is desirable to develop more accurate and automatic VQA tools that can be used to optimize video systems, towards providing satisfactory levels of quality of experience (QoE) to the end user [2].

To achieve optimal video quality under limited bandwidth, storage, and power consumption conditions, video encoding technologies commonly employ lossy coding schemes, which can cause compression artifacts that degrade perceptual quality [3]. Videos can also be degraded by transmission distortions (e.g., packet loss, playback interruption, and freezing) due to channel throughput fluctuations [4]. Hence, videos suffer not only from spatial distortions such as blocking, blurring, ringing, mosaic patterns, and noise, but also from temporal distortions such as motion compensation mismatches, flicker, mosquito effects, ghosting, jerkiness, smearing, and so forth [3].

Specifically, local flicker denotes the temporal fluctuation of spatially local luminance or chrominance in videos. Local flicker occurs mainly due to coarse quantization, mismatching of inter-frame blocks, improper deinterlacing, and dynamic rate changes in adaptive rate control [5]. Local flicker distortions, which are not well explained by current VQA models, frequently appear near moving edges and textures in compressed videos as well as in interlaced videos, producing annoying visual artifacts such as line crawling, interline flicker, and edge flicker [6], [7].

Since humans are the ultimate arbiters of received videos, understanding how humans perceive visual distortions and modeling the visibility of distortions in digital videos have been important topics for developing successful quality assessment models [8]. Early human visual system (HVS) based VQA models include Mannos and Sakrison’s metric [9], the Visual Differences Predictor (VDP) [10], Sarnoff Just Noticeable Differences (JND) Vision Model [11], Moving Pictures Quality Metric (MPQM) [12], the Perceptual Distortion Metric (PDM) [13], and the Digital Video Quality (DVQ) model [14]. Later models include Structural Similarity (SSIM) [15], Multiscale-SSIM (MS-SSIM) [16], motion-based SSIM [17], Visual Information Fidelity (VIF) [18], Visual Signal-to-Noise Ratio (VSNR) [19], Video Quality Metric (VQM) [20] and the Scalable Wavelet Based Video Distortion Index [21]. More recently, Ninassi *et al.* [22], TeraVQM [23], MOVIE [24], SpatioTemporal-Most Apparent Distortion (ST-MAD) [25], SpatioTemporal Reduced Reference Entropic Differences (STRRED) [26], Video-BLind Image Integrity Notator using DCT-Statistics (V-BLIINDS) [27], and VQM-Variable Frame Delays (VQM-VFD) [28] are examples that include more sophisticated temporal aspects. In video streaming services, other factors impact the overall QoE such as initial loading delays, freezing, stalling, skipping, and video bitrate, all of which have been widely studied [29]-[31].

One potentially important aspect of the design of VQA models that remains poorly understood is the effect of temporal visual masking on the visibility of temporal distortions. The mere presence of spatial, temporal, or spatiotemporal distortions does not imply a corresponding degree of perceptual

E-mail address: larkkwonchoi@gmail.com (L. K. Choi), bovik@ece.utexas.edu (A. C. Bovik).