Signal Processing: Image Communication 47 (2016) 426–437

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Objective and subjective evaluation of High Dynamic Range video compression

R. Mukherjee [a,*], K. Debattista [a], T. Bashford-Rogers [a], P. Vangorp [b], R. Mantiuk [c], M. Bessa [d], B. Waterfield [e], A. Chalmers [a]

[a] WMG, University of Warwick, United Kingdom
[b] Bangor University, United Kingdom
[c] University of Cambridge, United Kingdom
[d] UTAD, Portugal
[e] Jaguar Land Rover Automotive PLC, United Kingdom

ARTICLE INFO

Article history:
Received 11 May 2016
Received in revised form
1 August 2016
Accepted 1 August 2016
Available online 3 August 2016

Keywords:
HDR video
Compression algorithm
Quality assessment
Ranking
Rate-distortion

ABSTRACT

A number of High Dynamic Range (HDR) video compression algorithms proposed to date have either been developed in isolation or only-partially compared with each other. Previous evaluations were conducted using quality assessment error metrics, which for the most part were developed for qualitative assessment of Low Dynamic Range (LDR) videos. This paper presents a comprehensive objective and subjective evaluation conducted with six published HDR video compression algorithms. The objective evaluation was undertaken on a large set of 39 HDR video sequences using seven numerical error metrics namely: PSNR, logPSNR, puPSNR, puSSIM, Weber MSE, HDR-VDP and HDR-VQM. The subjective evaluation involved six short-listed sequences and two ranking-based subjective experiments with hidden reference at two different output bitrates with 32 participants each, who were tasked to rank distorted HDR video footage compared to an uncompressed version of the same footage. Results suggest a strong correlation between the objective and subjective evaluation. Also, non-backward compatible compression algorithms appear to perform better at lower output bit rates than backward compatible algorithms across the settings used in this evaluation.

1. Introduction

High Dynamic Range (HDR) imaging is able to capture, store, transmit and display the full range of real-world lighting with much higher precision than mainstream Low Dynamic Range (LDR) (also known as Standard Dynamic Range, SDR) imaging. However, a significantly large amount of data needs to be stored and processed for HDR video. To allow HDR video to be handled practically, a number of pre/post-processing (compression) algorithms have been proposed to convert HDR video data to an encoder suitable format. However, to date, these compression algorithms have only partially been compared with each other. This paper undertakes a comprehensive objective and subjective comparison of six previously published or patented HDR video compression algorithms and in doing so, follows a detailed methodology for evaluation and qualitative assessment of compressed HDR video content. In addition, a correlation is computed between the subjective and objective results for a better understanding of the shortcomings of current objective evaluation techniques for HDR video quality.

The primary contributions of this work are: (a) an objective evaluation of six HDR video compression algorithms using seven full-reference quality assessment (QA) metrics, (b) two subjective evaluations of the compression algorithms at two different output bitrates, using a ranking method with hidden reference conducted with 32 participants each; and (c) an assessment of the correlation between the objective and subjective evaluations.

In addition, the objective QA results are averaged over 39 sequences at 11 different quality settings to generalize the overall rate-distortion (RD) characteristics of the compression algorithms.

2. Related work

The acquisition of HDR imagery results in a large amount of floating point data which needs to be compressed in order to be handled efficiently on existing image or video infrastructure.

* Corresponding author.
E-mail address: ratnajitmukherjee@gmail.com (R. Mukherjee).

http://dx.doi.org/10.1016/j.image.2016.08.001
0923-5965/

This problem has been tackled through a variety of image and video compression algorithms which convert input HDR data to video encoder input formats. A brief review of compression algorithms is available in Banterle et al. [7].

Substantial research has also been conducted on evaluation of QA metrics for LDR image/video. Avcıbaş et al. [1] and Sheikh et al. [41] evaluated a number of QA metrics on distorted still images and concluded that metrics based on spectral magnitude error, perception, absolute norm and edge stability are most suitable for detecting image artefacts. They also conclude that although multiple QA metrics perform well on multiple image datasets none of them performed at par with subjective quality assessment. Seshadrinathan et al. [39] conducted an objective and subjective Video Quality Assessment (VQA) and concluded that dedicated VQA metrics such as Motion Based Video Integrity Evaluation (MOVIE) perform significantly better and have higher correlation with subjective results than still-image QA metrics.

In comparison, substantially less research has been conducted on development and evaluation of dedicated HDR QA metrics. Existing LDR QA metrics such as PSNR, SSIM [46] and VIF [40] have been extended to handle HDR values using Perceptually Uniform (PU) encoding [3]. Recently, a few full-reference HDR QA/VQA metrics have been proposed such as High Dynamic Range-Visible Difference Predictor (HDR-VDP) [28,34], High Dynamic Range-Visual Quality Metric (HDR-VQM) [36] and Dynamic Range Independent-Visual Quality Metric (DRI-VQM) [2].

A few evaluations have been conducted to test the performance of the proposed metrics. Čadík et al. [10] conducted an evaluation of HDR-VQA metrics with a dataset consisting of six HDR sequences using an HDR display and concluded that although the predictions by DRI-VQM and HDR-VDP are most suited for HDR-HDR image pairs, executing DRI-VQM becomes prohibitively expensive for sequences with greater than VGA resolution. Azimi et al. [4] tested the correlation between seven QA metrics and subjective quality scores with a dataset of 40 HDR video sequences and five types of distortions. The work demonstrates that HDR-VDP-2 [28] outperforms all other QA metrics when measuring compression induced distortions and has the highest correlation with the subjective quality scores. However, VIF [40] using PU encoding produces the best overall (tested against all distortions) results. Similar benchmarking evaluations of QA metrics for HDR image/video content have been conducted by Valenzise et al. [45], Mantel et al. [26] and Hanhart et al. [15] and Minoo et al. [32].

2.1. Evaluation of HDR video compression methods

Despite the research conducted into development and evaluation of QA/VQA metrics for both LDR and HDR content, very little has been done to evaluate existing HDR video compression algorithms using both QA metrics and psychophysical experiments. Koz and Dufaux [20] conducted a comparative survey on HDR video compression which compares the two different approaches to HDR video compression as explained later in Section 3.1. However, this work does not bring together objective and subjective evaluation techniques in order to provide a comprehensive evaluation of individual algorithms across a large set of sequences.

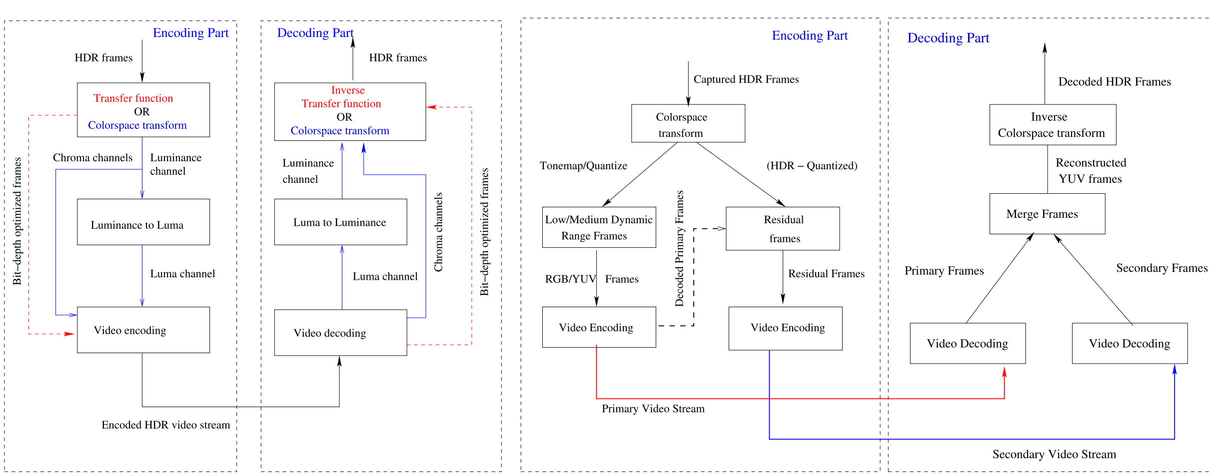
Recently, Hanhart et al. [17] conducted an evaluation of nine HDR video compression algorithms submitted in response to the Motion Pictures Experts Group (MPEG) committee's Call for Evidence (CfE) [25] to evaluate the feasibility of supporting HDR and Wide Color Gamut (WCG) content using the High Efficiency Video Codec (HEVC) [43] encoder. The paper concludes that the proposals submitted to MPEG can noticeably improve the standard HDR video coding technology and QA metrics such as PSNR-DE1000, HDR-VDP2 and PSNR-Lx can reliably detect visible difference. Azimi et al. [5] conducted a study to evaluate the compression efficiency of two possible HDR video encoding schemes (as defined in MPEG CfE [25]) based on the perceptual quantization of HDR video content [31] and tone mapping-inverse tone mapping with metadata. The paper concludes that for specific bitrates, subjective evaluation results suggest that HDR video generated by the perceptual quantization scheme were rated higher than the videos reconstructed using the inverse tone-mapping scheme. Similar evaluations on HDR video content have been conducted by Banitalebi-Dehkrodi et al. [6], Dong et al. [11], Rerabek et al. [38], Hanhart et al. [16], Narwaria et al. [35].

3. Method and materials

This section introduces the compression algorithms, sequences and overall research method followed for preparing the materials for the objective and subjective evaluation. The individual aspects of the objective and subjective evaluations are presented in Sections 4 and 5, respectively.

3.1. HDR video compression algorithms

HDR video compression algorithms can be classified into two approaches: *non-backward compatible* and *backward compatible*.



(a) *non-backward compatible* HDR video compression algorithm approach (b) *backward compatible* HDR video encoding approach

Fig. 1. Two generic approaches to HDR video encoding and decoding.