Signal Processing: *Image Communication* 39 (2015) 355–368

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Color difference weighted adaptive residual preprocessing using perceptual modeling for video compression ☆

Mark Q. Shaw [a,*], Jan P. Allebach [b], Edward J. Delp [b]

[a] Hewlett Packard, Boise, ID and School of Electrical and Computer Engineering, Purdue University, West Lafayette 47907, IN, USA
[b] School of Electrical and Computer Engineering, Purdue University, West Lafayette 47907, IN, USA

ARTICLE INFO

Available online 25 April 2015

Keywords:
Video coding
Perceptual modeling
Color appearance model
Human visual system (HVS)
Scalable video compression
Just noticeable distortion

ABSTRACT

In this paper, we investigate a method for selectively modifying a video stream using a color contrast sensitivity model based on the human visual system. The model identifies regions of high variance with frame-to-frame differences that are visually imperceptible to a human observer. The model is based on the CIE LAB color appearance model and the CIE ΔE_{94} color difference formula, taking advantage of the nature of frame-based progressive video coding. The use of a color contrast sensitivity model alone was not sufficient. Therefore, it was important to also incorporate perceptual saliency and spatial activity information from the scene. The method has been implemented in the JM 18.0 H.264/AVC encoder reference software and resulted in an average gain of 37% in data compression without perceptible degradation of the video quality. As expected, the amount of compression improvement obtained is dependent on the type of video content being compressed. The imperceptibility of the changes was confirmed through psychophysical evaluation.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Compression of video data is becoming increasingly essential in today's information age. Current video usage statistics [3] indicate that over 4 billion hours of video are watched each month on YouTube® alone. This does not include other video streaming services such as NetFlix®, Hulu® or Amazon Instant Video®. Video content is growing at an alarming rate, and the bandwidth requirements to support such video content are staggering.

Standardized video coding frameworks have been developed by the International Telecommunication Union (ITU) and the Motion Picture Experts Group (MPEG) [4–10]. Within these frameworks, the structure of the decoder is clearly defined, but the method by which the encoder creates the video stream is left to the discretion of the author of the encoder [11].

A variety of methods have been proposed to incorporate the effect of the human visual system (HVS) into video coding. Wu and Rao [12] have published a book that covers many aspects of perceptual coding, with a section dedicated to the use of computational models for Just Noticeable Differences (JND's) in the pixel domain. Of these approaches Chou et al. [13,14] have developed perceptually tuned subband image coders based on JND profiles, and Gu [15] developed a wavelet based video codec using a combination of Chou [13,14] and Jayant's [16] JND model. The approach of using preprocessing of the video sequence has been explored in [17–21]. Leung and Taubman [17] optimized the compression based on visual masking. In [22] and [23], Kim et al. and Fornesca and Ramirez tested the performance of various color difference models for moving images. Zheng et al. [24] developed a method that

☆ This work was presented in part at the 20th Color Imaging Conference [1] in 2012 and GlobalSIP 2014 [2].

* Corresponding author.

http://dx.doi.org/10.1016/j.image.2015.04.008
0923-5965/© 2015 Elsevier B.V. All rights reserved.

focused on spatial frequency sensitivity, Watson et al. [25] developed a video quality metric (DVQ) to compare two video sequences using a model that incorporates aspects of visual processing. Yang et al. [18,19] published papers on residual preprocessing based on a Just Noticeable Distortion (JND) profile. In these papers, the focus was on using a non-linear additive model for masking to enable a more accurate JND calculation. Chen et al. [26] demonstrated compression gains of 18-20% using color JND residue pre-processing in the DCT domain. Zhang et al. [27] proposed a foveation weight model based on auto-stereoscopic display, and incorporated it into a stereoscopic JND model with average bit-rate reduction gains of 5.7%. Chen and Guillemot [28] developed a foveated JND model that takes into account the foveation properties of the HVS, and Naccari [29] proposed a JND model based on spatio-temporal HVS masking mechanisms, which they claimed outperforms the work of Chen [28]. More recently, Liu [30] published a paper on the use of spatio-temporal JND profiles to process the motion-compensated residues to achieve a lower bit rate or reconstruction accuracy.

Lee and Ebrahimi [31] wrote a comprehensive summary of the advances in the understanding of perceptual video compression, reviewing recent progress in three main areas: perceptual modeling, implementation, and performance evaluation. Their work highlights the key principles that have been used in the various coding techniques.

The majority of the human vision perception mechanisms attempt to account for factors such as contrast sensitivity, visual masking, foveation, and visual attention. In all of the JND-based models that were reviewed by the authors of this paper, the predominant motivation for the use of the JND metric was attributed to the Weber–Fechner law dating back to the 1800s. Since then, there has been a significant amount of work on understanding the response of the HVS to differences in color, going beyond the impact of Lightness alone. The choice of the color space in which to perform the masking is very important; and the threshold of perceptibility for changes in the YUV color space is known to not be constant throughout the color space. It is extremely important that any JND measure not only take into account the luminance differences, but also the impact of chroma on an observers' ability to discriminate color differences [26,32–34].

This paper investigates a method to selectively discard inter-frame differences based on underlying assumptions about the sensitivity of the HVS to color difference. By taking into account the variation of the sensitivity of the HVS as a function of hue, chroma, and lightness [35], the method proposed herein selectively attenuates inter-frame differences based on a variance-weighted chromatic activity map. Inter-frame video coding takes advantage of the fact that not every pixel within a video sequence may change significantly from one frame to the next. By removing the redundancy of unchanging pixels, the video stream will only code those pixels that are changing from frame to frame, resulting in a significant improvement in the bit rate. This work uses a new image domain metric, based on principles of the HVS, visual saliency, and spatial activity of the frame for color video that takes into account differences in color perception, not just luminance. One of the underlying assumptions is that the pixel differences to be encoded from frame to frame should be perceptually significant. The algorithm has been implemented in the JM 18.0 H.264/AVC encoder reference software to demonstrate the impact of the residual processing for a constant q-factor H.264 encoder tested on a variety of CIF and HD video sequences.

The performance of our method is evaluated by comparing the file size of the compressed constant q-factor bit streams with and without our method, and by subjective evaluation of the compressed video sequences. The method is found to yield an average of 14% improvement in compression without visibly degrading the video quality. Further compression gains (averaging up to 37%) are achievable if one dynamically changes the color difference attenuation allowed in the encoding process. As expected, the amount of compression improvement obtained is dependent on the type of video content being compressed and the configuration of the video encoder.

In [1], we demonstrated our concept of residual preprocessing using an MPEG 2 simulation. In [2], we implemented the algorithm in an MPEG 4 framework, added a dynamic tone map model based on visual saliency, added chroma residue pre-processing, and performed a full psychophysical evaluation of a selection of compressed video streams. In this paper, we present a more complete exposition of these concepts, and present for the first time, results of applying our algorithm to HD video streams.

The paper is organized as follows. In Section 2, we summarize the fundamental principles of the HVS. In Sections 3.1 and 3.2, we describe how the perceptibility map and the spatial activity maps are determined for the current frame being encoded. In Section 3, we outline how the key components of the proposed workflow, the color difference map and spatial activity map, are used to attenuate the residuals that will be encoded. In Section 4.1, we summarize the implementation in the JM 18.0 H.264/AVC encoder reference software. Finally, in Section 4.2, we outline the experimental procedures and the results that demonstrate the performance of the algorithm.

2. Principles of human vision

In this section, we review some basic principles concerning the HVS to provide a basis for our perceptual model. Research in the area of color science has shown that not all color differences are equally likely to be perceptible. The HVS sensitivity to changes in color varies as a function of Lightness, Chroma, and Hue [33]. Fig. 1 illustrates this phenomenon, using the 1931 CIE chromaticity diagram which is bounded by the spectral locus of pure colors. Here, the enlarged ellipses represent loci of colors that were perceived by an observer as having a single JND from the color at the center of the ellipse. If the color space were perceptually uniform, one would expect the ellipses to be circles of constant radius throughout the color gamut of the device. It is evident that the ellipses in the top-center (green region of chromaticity diagram) are much larger than those in the bottom left (blue region). This indicates that an observer is much more sensitive to color changes in the blues, than in the greens.

Download English Version:

https://daneshyari.com/en/article/537183

Download Persian Version:

https://daneshyari.com/article/537183

Daneshyari.com