Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Augusst 2016, 23(4): 77–82
www.sciencedirect.com/science/journal/10058885
The Journal of China Universities of Posts and Telecommunications
http://jcupt.bupt.edu.cn

Reduced frame set on wireless distorted video for quality assessment

Shao Hua (✉), Wen Xiangming, Lu Zhaoming, Chen Yawen, Lu Jingyu

1. Beijing Laboratory of Advanced Information Networks, Beijing University of Posts and Telecommunications, Beijing 100876, China
2. Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract

Objective video quality assessment methods often evaluate all the frames regardless of their importance. For wireless distorted videos, not every frame has the same contribution to the final overall quality due to the channel fading and interference, which may lead to the capacity variation in temporal. Besides, with the content similarity and error propagation pattern in temporal domain, it is possible to evaluate the overall quality with only part of the frames. In this paper, a demonstration is performed to show that the video quality can be evaluated with reduced frames set (RFS), and a state transition model is proposed to extract the RFS. At last, a video quality assessment (VQA) method is carried out based on RFS. Compared with several state-of-the-art methods, our method can achieve a suitable accuracy with less frames to be processed.

Keywords VQA, wireless distortion, quality of experience

1 Introduction

Both multimedia technology and mobile communications have experienced massive growth and commercial success in recent years. One main aim of designing visual communication systems is to use the least resources to achieve the highest visual quality with respect to certain constraints such as bit rate, complexity, and maximum delay [1]. Evaluating the perceptual quality of video is of tremendous importance in design and optimization of wireless video processing and transmission systems [2]. Different from the traditional video coding system, video transmission over wireless channel with good quality or low end-to-end distortion is particularly challenging since the received video is subject to not only quantization errors but also transmission distortion. Compared with quantization errors, transmission distortion caused by packet loss during the transmission is the major part of the end-to-end distortion in wireless video communication [3].

A packet transmitted over a wireless channel is susceptible to bit errors due to attenuation, shadowing, fading and multiuser interference. With the smaller packet sizes in wireless applications, the observed artifacts are spatiotemporally localized and appear different from the artifacts observed in Internet protocol (IP) networks [4]. With the content and impairments in nearby frames sharing the similar spatial and temporal character, there is much redundancy content to be computed when evaluating in the traditional way. Besides, evidence has shown that the severe distortion has a temporal hysteresis effect on the overall results [5], and according to the research of human visual system (HVS), the overall quality of a video is seriously affected by the quality variations [5–8].

Many state-of-the-art objective video quality methods focus on extracting quality-related features from the videos in spatial or temporal domain, and a pooling process is performed to obtain an overall quality value [9–10]. Motion-based video integrity evaluation index (MOVIE) in Ref. [11] uses 105 filters to capture the spatial and temporal distortion every eight frames to achieve the overall quality. Structural similarity (SSIM) [12] computes the every-frame structural similarity and an overall quality is obtained by pooling all these scores. Seldom work has

Received date: 05-05-2016
Corresponding author: Shao Hua, E-mail: sarathy@bupt.edu.cn
DOI: 10.1016/S1005-8885(16)60048-1

been done to explore the temporal distortion redundancy and quality importance of frames. For the wireless distorted videos, the number of frames which suffered from severe packet loss and would seriously degrade the quality may be limited. The rest distortion may be caused by error propagation, motion estimation mismatch etc., and due to the similarity of neighboring frames, the contribution of the rest frames may be limited or estimated. The overall video quality may be largely determined by just a few number of importance frames.

In this paper, we focus on wireless distorted videos and are seeking to find a set of frames which can be used to evaluate the overall objective video quality. These frames are denoted as the RFS for VQA in this paper. The main contributions of this paper are two-fold:

1) A demonstration is conducted to explore the RFS in video and a state transition method is proposed to extract the RFS. Analysis indicates that it is feasible to evaluate the objective video quality with just RFS.

2) VQA model is provided based on RFS. Testing results show the method achieves a suitable accuracy with less frames to be computed.

The demonstration to explore the existence and properties of RFS in distorted video is conducted in Sect. 2. A method to extract the RFS from the video based on a state transition model is also presented in Sect. 2. Sect. 3 presents an objective video quality model and Sect. 4 presents its performance. Sect. 5 concludes the paper.

2 Reduced frame set extracting method

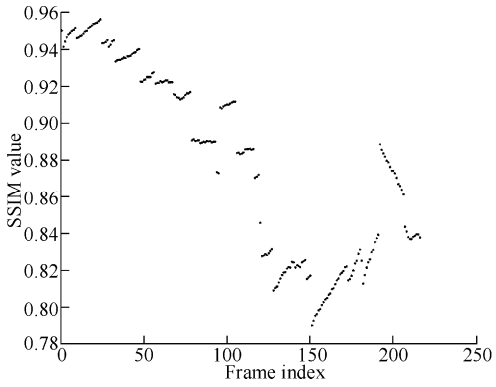
2.1 Reduced frame set for VQA

In this section, a demonstration based on two well-known image quality metric (IQM) is conducted to explore the RFS. The laboratory for image and video engineering (LIVE) video database [4] is used as the dataset and the frame quality is evaluated one by one in order to show the temporal character.

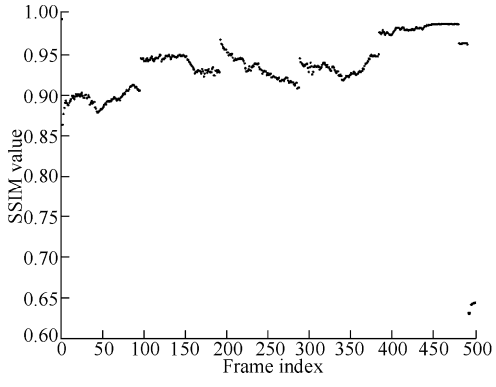
LIVE is a rich and powerful video database for VQA test. It contains ten uncompressed, high-quality, source videos of natural scenes in resolution of 768×432 YUV 4:2:0 format. All videos are about 10 s. fifteen test sequences from each of the source sequence using four different processes of distortions, i.e. motion picture experts group-2 (MPEG-2) compression, H.264 compression, simulated transmission of H.264 compressed bit-stream through error-prone IP networks, and through error-prone wireless networks are provided. In this demonstration all the wireless distorted sequences (40 videos in total) from LIVE are used as the dataset.

SSIM index [12] is a widely used image quality assessment (IQA) metric proposed by Wang et al. with the philosophy that HVS is evolved to figure out structure impairments. For its many good prosperities and specific physical meaning, SSIM index has been widely recognized as a benchmark for IQA. Visual information fidelity (VIF) [13] is proposed by Sheikh et al. by approaching the IQA problem as an information fidelity problem. It quantifies the information that could ideally be extracted by the brain from the reference image. SSIM and VIF are based on two different standpoints and are widely used by researchers as criteria. It is noted that not only SSIM and VIF but also any other IQA metric can be used to get a similar result.

For each distorted sequence in the dataset, both SSIM and VIF are ran frame by frame. The quality of frames are plotted in time sequence to visualize its temporal properties. Due to the page limitation, only ‘blue sky’ (bs2_25fps) and ‘park run’ (pr2_50fps) are illustrated in Fig. 1.



(a) SSIM of ‘bs2_25fps’



(b) SSIM of ‘pr2_50fps’