INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 13, Number 6, Pages 879–897

MATHEMATICAL MODELS FOR QUALITY ANALYSIS OF MOBILE VIDEO

SONGQING ZHAO, HONG JIANG, CHAO LIANG, SHERIF SHERIF, AND AHMED TARRAF

Abstract. With the explosive growth of mobile video applications, analysis of video quality becomes increasingly important because it is an important Key Performance Indicator (KPI) for Quality of Experience (QoE). In this paper, a framework for non-reference video quality analysis is proposed and applied to Video Telephony (VT) in LTE networks. Three metrics, blockiness, blur and freezing, are used to estimate the MOS. Blockiness is detected by taking the H.264 codec features into account, blur is estimated by utilizing the percentage of noticeable blurred edges in each frame, and freezing is evaluated by using a sigmoid function to mimic the effect of different freezing duration on the Human Visual System (HVS). Furthermore, the three metrics are combined into one objective MOS by considering different weighting factors and using the linear curve fitting. Above 90% correlation is achieved between the objective MOS score and subjective MOS score.

Key words. Mathematical models, video quality analysis, quality of experience, mean opinion score, blockiness, blur, freezing, human visual system, modeling.

1. Introduction

As smartphones and tablets are increasingly being used by more and more people and cellular network capacity is being significantly improved with more 4G LTE network deployment, mobile data traffic is growing explosively. Among the data traffic, video traffic is playing a big role. In a recent study [5], mobile video traffic was already 51 percent of the entire mobile data traffic by the end of 2012. It forecasts that mobile video will increase 16 times between 2012 and 2017 and two-thirds of the world's mobile data traffic will be video by 2017. As a consequence the analysis of video quality is becoming increasingly important because it is an important Key Performance Indicator (KPI) for Quality of Experience (QoE). In recent years, QoE research and development has gained significant attraction in both academia and industry since the first international workshop on Quality of Multimedia Experience (QoMEX) was held in 2009 and more and more video quality models are recommended by Video Quality Expert Group (VQEG) to the ITU-T standardization process [25].

Fig. 1 is a typical architecture for Video Telephony (VT) over all-IP based LTE network. The User Equipment (UE) of the originator sends out the original video through the uplink indicated by blue lines, and the UE terminator receives the video through the downlink indicated by red lines. The real-time video for VT application is usually encoded in H.264 and transmitted through the Real-time Transport Protocol (RTP). During transmission, especially over the air interface between UE and base station, IP packets may experience network impairments such as packet loss, delay, and jitter, which are crucial factors in causing degradation to the quality of the received video. Video packets may be lost due to network congestion or they may be discarded by the video decoder if they arrive at the

Received by the editors July 22, 2016.
2000 *Mathematics Subject Classification.* 35R35, 49J40, 60G40.

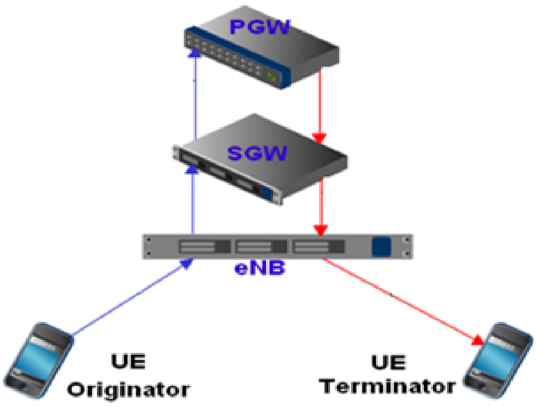

FIGURE 1. Components in LTE network.

terminator UE with a delay so large that it exceeds the video de-jitter buffer's limit. When the terminator UE does not receive the video packets for certain time duration, it will use the latest decoded frame for display, resulting in freezing or jerky video. When a few video packets in one Group of Picture (GOP) are lost, the video decoder will produce some impaired frames even after error concealment, which cause annoying blockiness and blur. Once there is an impaired frame, the subsequent frames until the next I frame will be also adversely affected due to the error propagation property in H.264 codec. It is worthwhile to mention that although conceptually the blockiness occurs on the regular 8x8 block boundaries, and has clear block edges, this notable feature is diminished due to the de-blocking filter and error concealment at the video decoder, making blockiness detection more challenging.

There exist subjective video quality and objective video quality. Subjective video quality can be evaluated according to the standard P.910 [11]. Essentially, a group of people are asked to watch the video in a specific environment with certain lighting requirements and to give their scores to the video, according to their liking, using a certain scale, e.g. 1-5. After averaging the scores over the audience, each video is provided with a Mean Opinion Score (MOS). However, this process to obtain the subjective MOS is both time and resource consuming, especially when a lot of video sequences need to be evaluated. The purpose of objective video quality metrics [23] is to use artificial intelligence to replace people evaluating the video. Such a predicted MOS approximates the subjective MOS by detecting and estimating the impairments most sensitive to the Human Visual System (HVS), e.g. freezing/jerkiness, blockiness, and blur, and combining these impairments into one MOS.

In general, there are three categories for objective video quality analysis. The first category is Full Reference (FR) video quality, in which the analysis is performed by comparing the received and decoded video with the original reference video. The most used FR metric is Peak Signal-to-Noise Ratio (PSNR) because of its computational simplicity and simple mathematical formula. However, it correlates poorly with the subjective ratings [3] [22]. Thus many new FR metrics are developed that better correlate with subjective ratings than PSNR does, such as structure similarity (SSIM) [22] and visual signal-to-noise ratio (VSNR) [3], and the latest FR video quality standard for multimedia application is ITU-T J.247 [10]. Another