# REDUCED-REFERENCE PERCEPTUAL QUALITY ASSESSMENT FOR VIDEO STREAMING

*Mohammed A. Aabed and Ghassan AlRegib*

Center for Signal and Information Processing (CSIP)
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332, U.S.A.
{maabed, alregib}@gatech.edu

## ABSTRACT

We propose a perceptual video quality monitoring metric for streaming applications using the optical flow features. This approach is a reduced-reference pixel-based and relies only on the deviation of the optical flow of the corrupted frames. This techniques compares an optical flow descriptor from the corrupted frame against the descriptor obtained from the anchor frame. This approach is suitable for videos with complex motion patterns. Our technique does not make any assumptions on the coding conditions, network loss patterns or error concealment techniques. We validate our proposed metric by testing it on a variety of distorted sequences from the LIVE mobile database. Our results show that our metric estimates the perceptual quality at the sequence level accurately. We report the correlation coefficients with the differential mean opinion scores (DMOS) reported in the database. The results show Spearman's correlations of 0.88 for the tested sequences.

***Index Terms***— video quality monitoring, channel distortion, video streaming, HEVC, H.264/MPEG-4 AVC, optical flow, network losses

## 1. INTRODUCTION

The continuous growth of mobile Internet traffic in general, and video traffic in particular, has triggered the communication's community concern with bandwidth and quality of experience (QoE). Global IP traffic has increased fivefold over the past 5 years. It also is predicted to reach 1.6 zettabytes by 2018 (threefold increase from 2013), out of which 79% will be video traffic (66% in 2013). A million minutes of video content is estimated to cross the network every second by 2018. The Global and North America (NA) Mobile Data Traffic Forecast Update [1] released earlier this year estimates a growth of 45.7% in global mobile connections reaching 10.2 billion mobile connections. Mobile video traffic will account for over 69% of that total. Furthermore, busy-hour Internet traffic is growing more rapidly than average Internet traffic [2]. Thus, the standardization bodies are adapting to this growth by motivating technologies that increase the efficiency of bandwidth utilization, data compression and quality of experience (QoE). All of these issues and others open new challenges in quality assessment, error concealment, etc. This paper analyzes the impact of network losses on the fidelity of the decoded video and proposes a new approach to measure channel-induced distortion.

The problem of quality assessment for streamed video sequences has been addressed in several papers in the literature [3, 4, 5, 6, 7, 8]. In [3], the performance of subjective quality assessment campaign of the HEVC standard involving 494 test subjects was reported. The performance of MS-SSIM and the General VQM for high-definition videos is investigated in [6]. It is shown that MS-SSIM and VQM outperform PSNR on both HD and non-HD data and that MS-SSIM is slightly better on all databases. Hanhart *et al.* [4] tested the performance of various full-reference (FR) quality metrics on 4k UHD videos. PSNR, VSNR, SSIM, MS-SSIM, VIF, and VQM metrics were accurate in distinguishing different quality levels for the same content. Konuk *et al.* [7] used motion vectors, bit rate, and packet loss ratio to propose a video quality assessment algorithm. In [5], the authors combine subjective tests and objective analysis to propose a video quality assessment method. The paper uses a subject-objective mapping strategy with four indicators that affect video quality to set the relationship between the indicator and user experience. McLaughlin and Hemami [8] have shown that video quality due to packet loss can be estimated at the decoder after concealment using a proposed reduced-reference (RR) approach that uses averages across superblocks of four macroblock parameters, along with the received motion vectors. The work herein addresses the objective perceptual quality assessment of streamed videos subject to network losses with access only to the decoded videos.

In this paper, we propose a RR optical flow based video quality assessment approach for streamed videos. We intro-

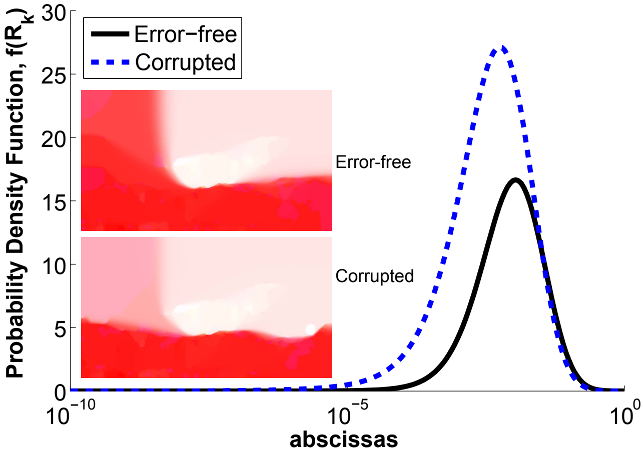

**Fig. 1**. The empirical probability density functions of $\mathbf{R}_{240}$ of the corrupted and error-free `Landing Airplane` sequences. The SSIM value for the distorted frame is 0.986.

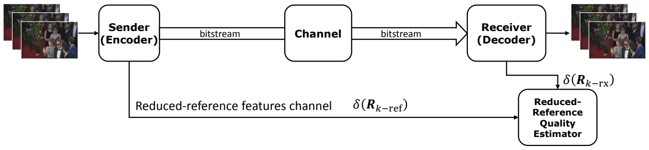

**Fig. 2**. Reduced-reference quality estimation configuration.

duce a perceptual quality metric based on analyzing and fusing the statistical features of the optical flow in the received and anchor videos. The decoder calculates a descriptor of the optical flow map of the received frame and compares it against the descriptor obtained from the reference frame. This approach does not make any assumptions on the coding standard or parameters, concealment technique or network conditions. It blindly operates on the decoded video after the decoder. We argue that the change in the optical flow due to channel-induced losses can be used to capture the distortion in the frames, and subsequently at the sequence level.

The rest of this paper is organized as follows. In Section 2, we explain our approach to measure the channel induced distortions, which utilizes the statistical features of the optical flow. Section 3 details the simulations setup and test sequences used in the experiments herein, followed by the results and analysis of the model validation experiments. Finally, Section 4 concludes the paper and outlines future directions of this work.

## 2. OPTICAL FLOW BASED PERCEPTUAL QUALITY ASSESSMENT

In this section, we explain our proposed methodology for video quality assessment using the optical flow. The proposed approach is based on the fact that any channel induced distortion will cause a temporal inconsistency in the optical flow. The Human Visual System (HVS) observes the distortion in the form of visual discomfort due to inconsistency in the pixels or objects in the distorted frame.

**Algorithm 1** Iterative optical flow frame-level processing algorithm

```
Initialization:
$T(\mathbf{R_k}, 0) \leftarrow \mathbf{R_k}$
$\delta(\mathbf{R}_k, 0) \leftarrow \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} |T(\mathbf{R}_k, 0)|$
$t = 0$
while $\delta(\mathbf{R}_k, t) \leq$ threshold do
    $t = t + 1$
    $T(\mathbf{R}_k, t) = \downarrow_{\mathbf{2}} [T(\mathbf{R}_k, t-1) * \phi(\sigma)]$
    $\delta(\mathbf{R}_k, t) = \delta(\mathbf{R}_k, t-1) + \sum_{i=1}^{N-1} \sum_{j=0}^{M-1} |T(\mathbf{R}_k, t)|$
end while
$\delta(\mathbf{R}_k) = \delta(\mathbf{R}_k, t)$
return $\delta(\mathbf{R}_k)$
```

Fig. 1 shows the visualizations of the optical flow maps of the original frame and a distorted one, respectively. Furthermore, Fig. 1 shows the probability density functions (PDFs) of these optical flows, $\mathbf{R}_k$, for the error-free and corrupted frames. The figures show that there is a discrepancy in the optical flow due to distortion in the frame. It should be noted that the SSIM values of the corrupted frame in this case is 0.986.

In this paper, our goal is to capture these inconsistencies in the optical flows throughout the video due to the channel-induced losses. Let $f_k$ be the frame of interest. Furthermore, let $\mathbf{U}_k$ and $\mathbf{V}_k$ denote the matrices of its horizontal and vertical optical flow velocities, respectively. Furthermore, let $\mathbf{R}_k$ denote the matrix of magnitudes of the flow velocities [9]:

$$\mathbf{R}_k = \sqrt{\mathbf{U}_k^2 + \mathbf{V}_k^2} \tag{1}$$

where $k$ is the temporal index of the frame in the received video. All the results and experiments in this paper were obtained using the Horn-Schunck optical flow estimation method [9]. This approach, nevertheless, is valid for any optical flow estimation algorithm.

Algorithm 1 details the process performed to each optical flow map for every frame. The notations used in the Algorithm 1 are as follows: $t$ is the iteration index, $T(\mathbf{R}_k, t)$ is the downsampled version of $T(\mathbf{R}_k, t-1)$ at iteration $t$, $\downarrow_{\mathbf{2}} [\cdot]$ is a downsampling operator, and $\phi(\alpha)$ is a Gaussian kernel with standard deviation $\alpha$.

In an iterative manner, we look at the aggregate of magnitudes of the optical flow map at different scales. This approach was inspired by the diffusion distance dissimilarity metric [10]. The two expressions in (2) are the core of this iterative process. In every iteration, we downsample after smoothing the map with a Gaussian kernel. Then, we take the aggregate of the magnitudes of this filtered version of optical flow map as an output from each iteration. The aggregates from all the iterations are finally accumulated into one descriptor. This process yields a descriptor of an optical flow map, $\delta\left(\mathbf{R}_k\right)$.

$$T\left(\mathbf{R}_k, t\right) = \downarrow_{\mathbf{2}} \left[T\left(\mathbf{R}_k, t-1\right) * \phi\left(\sigma\right)\right] \tag{2a}$$

$$\delta\left(\mathbf{R}_k, t\right) = \delta\left(\mathbf{R}_k, t-1\right) + \sum_{i=1}^{N-1}\sum_{j=0}^{M-1}\left|T\left(\mathbf{R}_k, t\right)\right| \tag{2b}$$

In previous work [11], we used this algorithm in a no-reference frame-level configuration to estimate the perceptual quality of each distorted frame. Our goal in this paper is to explore the utility of this framework for quality monitoring at sequence level. The human visual system does not make an independent decision on the frames. It rather forms a decision on the quality of perceptual experience based on a collective evaluation of the frames. For an objective quality metric to capture the subjective quality accurately at the sequence level, the processing of the descriptors and pooling strategies have to be regulated and examined.

Assuming a reduced-reference quality estimation configuration, as depicted in Fig. 2, the reference descriptor $\delta\left(\mathbf{R}_{k-\text{ref}}\right)$, is estimated at the encoder and available at the decoder or quality monitor. The decoder performs an identical operation and calculates the descriptor for the received frame, $\delta\left(\mathbf{R}_{k-\text{rx}}\right)$. The difference between the descriptors from the reference and received videos are then used as a perceptual quality estimator as follows:

$$D_k = \left|\delta\left(\mathbf{R}_{k-\text{ref}}\right) - \delta\left(\mathbf{R}_{k-\text{rx}}\right)\right|. \tag{3}$$

This metric captures the inconsistencies in the optical flow map at the frame level. To estimate the perceptual quality at the sequence or GOP level, $P$, we simply calculate the arithmetic mean for the picture set of interest:

$$P = \log\left[\underset{\forall k}{\mathrm{E}}\left[D_k\right]\right]. \tag{4}$$

This expression represents a perceptual quality estimator for the received sequence at the sequence level. As we will show in Section 3, this process yields a predictor that correlates well with the DMOSs at the sequence level.

## 3. EXPERIMENTS AND RESULTS

To test and verify the proposed reduced-reference model, we used the LIVE Mobile Video Quality Assessment database [13]. These videos were coded using the H.264 scalable video codec (SVC) standard. The sequences have 450 frames and a resolution of $1280 \times 720$. The corrupted sequences are taken from the wireless channel packet-loss set in the database. The database contains various distorted versions of every sequences. We randomly selected at least two different distorted version of each sequence. The JM software implementation of the H.264 scalable video coding (SVC) was used to code all the sequences [14]. The videos were encoded with a GOP period of 16 and packet-loss rates set to 3% for the reference decoded sequences. The packet size was set to 200 bytes following the recommendation for wireless transmission. To introduce errors to the bitstream, the coded bitstream was transmitted over a simulated noisy wireless channel [13].

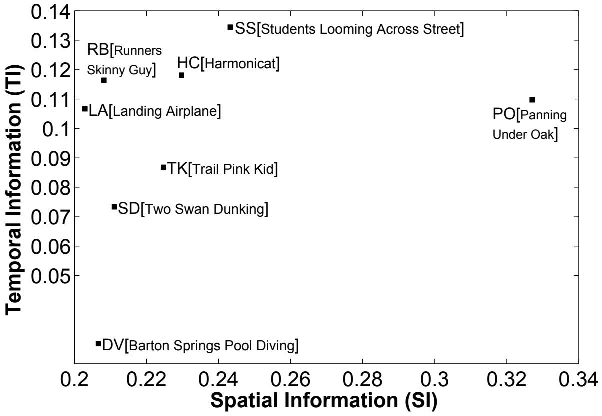

**Fig. 3**. Spatial information (SI) versus temporal information (TI) indices for the selected sequences [12].

The proposed metric was tested on a variety of video sequences from the LIVE mobile database. The selected videos span different temporal and spatial complexities. Fig. 3 shows the spatial information (SI) and temporal information (TI) indices on the luminance channel for the selected sequences, as per the recommendation in [12]. A high score on the SI and the TI scale indicates a higher spatial and temporal features of the test sequence.

### 3.1. Results and Analysis

Tables 1-2 show Pearson's Correlation Coefficients (PCC) and Spearman's Correlation Coefficient (SCC) between the temporally estimated perceptual quality at the frame level, $D_k$, and temporal Mean Opinion Score (MOS) reported in the database. For different distorted versions of every sequence, we report the median PCC and SCC values. Furthermore, the DMOS values reported in the Mobile LIVE database are calculated using percentile pooling using only the lowest $5\%$ MOS values. Hence, the correlations reported in Tables 1-2 are for these temporal MOS values.

Fig. 4 shows scatter plot for the estimated perceptual quality, $P$ (4), versus the DMOS values at the sequence level. This plot shows that the value of $P$ is highly correlated with the DMOS reported in the database [13]. The proposed reduced-reference metric captures the DMOS pattern as it can be seen

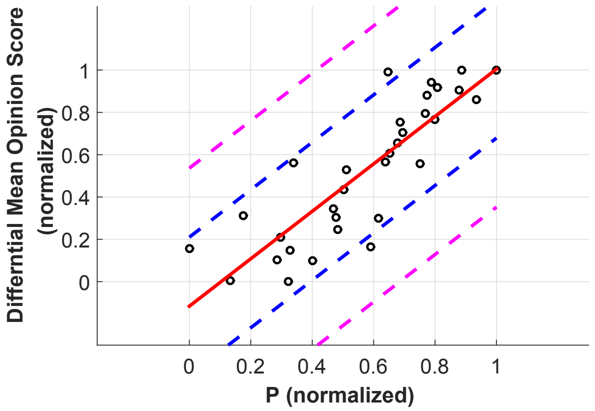

**Fig. 4**. The proposed reduced-reference perceptual quality metric scatter plot versus the reported DMOS for all the test sequences. The blue and pink lines are $P \pm \sigma$ and $P \pm 2\sigma$, respectively, where $\sigma$ is the data standard deviation.

by looking at the black points on the plot. We note that these results were obtained without performing any curve fitting.

For the purpose of measuring the accuracy of the estimated perceptual quality, we calculate Pearson's and Spearman's correlation coefficients between the estimated perceptual quality based on the proposed RR approach and the recorded DMOS of the corrupted sequences. Table 3 summarizes the results for all the tested sequences for PCC and SCC. The authors in [13] reported that the highest correlation coefficients for FR approaches with DMOSs were obtained using the Noise Quality Measure (NQM) metric [15] and Visual Information Fidelity (VIF) metric [16]. Hence, we report the correlation coefficients for these FR metrics at the sequences level for comparison. The proposed model correlates well with the DMOS values. The correlation coefficients for all test sequences are 0.85 and 0.88 for both Pearson's and Spearman's correlation coefficients, respectively. In particular, the proposed approach works well for the sequences with high temporal complexity such as the `Students Looming Across Street` and `Harmonicat` video sequences. Furthermore, our quality metric works well for sequences with medium or low temporal complexity, such as `Barton Springs Pool Diving` and `Two Swan Dunking`. These results show the effective utility of relying on the optical flow in analyzing the distortion in videos with complex motion patterns where the temporal feature across frames are highly variant.

## 4. CONCLUSION

We propose in this paper a reduced-reference perceptual video quality metric to estimate the channel-induced distortion due to network losses. The proposed technique does not make any assumption about the coding conditions or video sequence. It rather explores the temporal changes between the frames by analyzing the variations in the statistical properties of the optical flow. We validate our approach by testing it on various sequences and compare our estimated quality metric with the DMOS values at the sequences level reported in the LIVE mobile database for sequences subject to network errors or losses. Our experiments show that the proposed technique captures the perceptual quality very well.

**Table 1**. Pearson's correlation coefficients (PCC) with the reported temporal MOS.

| Sequences | Proposed $D_k$ (3)<br>(RR) | VIF [16]<br>(FR) |
|---|---|---|
| `Landing Airplane` | 0.92 | 0.92 |
| `Two Swan Dunking` | 0.91 | 0.92 |
| `Runners Skinny Guy` | 0.93 | 0.94 |
| `Students Looming Across St.` | 0.91 | 0.96 |
| `Panning Under Oak` | 0.92 | 0.95 |
| `Barton Springs Pool Diving` | 0.82 | 0.95 |
| `Trail Pink Kid` | 0.87 | 0.92 |
| `Harmonicat` | 0.93 | 0.90 |

**Table 2**. Spearman's correlation coefficients (SCC) with the reported temporal MOS.

| Sequences | Proposed $D_k$ (3)<br>(RR) | VIF [16]<br>(FR) |
|---|---|---|
| `Landing Airplane` | 0.98 | 0.94 |
| `Two Swan Dunking` | 0.98 | 0.95 |
| `Runners Skinny Guy` | 0.98 | 0.97 |
| `Students Looming Across St.` | 0.98 | 0.98 |
| `Panning Under Oak` | 0.99 | 0.94 |
| `Barton Springs Pool Diving` | 0.98 | 0.97 |
| `Trail Pink Kid` | 0.99 | 0.91 |
| `Harmonicat` | 0.99 | 0.86 |

**Table 3**. The overall sequence level PCC and SCC between the DMOS and the metrics.

| Sequences | Proposed, $P$ (4)<br>(RR) | NQM [15]<br>(FR) | VIF [16]<br>(FR) |
|---|---|---|---|
| Overall PCC | 0.8523 | 0.8738 | 0.8979 |
| Overall SCC | 0.8820 | 0.8985 | 0.8739 |

## 5. REFERENCES

[1] "Cisco visual networking index (vni) global and north america (na) mobile data traffic forecast update 2013-2018," http://graphics8.nytimes.com/packages/pdf/technology/cisco-mobile-forecast.pdf, Feb 2014, Accessed: 2014-12-21.

[2] "The zettabyte era: Trends and analysis," http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.pdf, Feb 2014, Accessed: 2014-12-21.

[3] Francesca De Simone, Lutz Goldmann, Jong-Seok Lee, and Touradj Ebrahimi, "Towards high efficiency video coding: Subjective evaluation of potential coding technologies," *J. Vis. Comun. Image Represent.*, vol. 22, no. 8, pp. 734–748, Nov. 2011.

[4] Philippe Hanhart, Pavel Korshunov, and Touradj Ebrahimi, "Benchmarking of quality metrics on ultra-high definition video sequences," in *2013 18th International Conference on Digital Signal Processing (DSP)*, 2013, pp. 1–8.

[5] Weitao Wang, Yuehui Jin, Tan Yang, and Yidong Cui, "A video quality assessment method using subjective and objective mapping stategy," in *Cloud Computing and Intelligent Systems (CCIS), 2012 IEEE 2nd International Conference on*, Oct 2012, vol. 02, pp. 514–518.

[6] S. Wulf and U. Zolzer, "Full-reference video quality assessment on high-definition video content," in *Signal Processing and Communication Systems (ICSPCS), 2012 6th International Conference on*, Dec 2012, pp. 1–10.

[7] Baris Konuk, Emin Zerman, Gokce Nur, and Gozde Bozdagi Akar, "A spatiotemporal no-reference video quality assessment model," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, Sept 2013, pp. 54–58.

[8] Linda McLaughlin and Sheila S. Hemami, "Reduced-reference quality assessment with scalable overhead for video with packet loss," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, Sept 2013, pp. 1622–1626.

[9] Berthold K.P. Horn and Brian G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–204, 1981.

[10] Haibin Ling and K. Okada, "Diffusion distance for histogram comparison," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, June 2006, vol. 1, pp. 246–253.

[11] M.A. Aabed and G. AlRegib, "No-reference perceptual quality assessment of streamed videos using optical flow features," in *Global Conference on Signal and Information Processing (GlobalSIP), 2014 IEEE*, Dec 2014.

[12] ITU-T, "P.910: Subjective video quality assessment methods for multimedia applications," Tech. Rep., ITU Telecommunication Standardization Sector, 2008.

[13] A.K. Moorthy, Lark Kwon Choi, A.C. Bovik, and G. de Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 652–671, Oct 2012.

[14] "Svc reference software (jsvm software), joint video team (jvt)," http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm, Accessed: 2014-06-01.

[15] N. Damera-Venkata, T.D. Kite, W.S. Geisler, B.L. Evans, and A.C. Bovik, "Image quality assessment based on a degradation model," *Image Processing, IEEE Transactions on*, vol. 9, no. 4, pp. 636–650, Apr 2000.

[16] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430–444, Feb 2006.