



HAL
open science

Evaluation of Video Quality Monitoring Based on Pre-computed Frame Distortions

Dominik Klein, Thomas Zinner, Kathrin Borchert, Stanislav Lange, Vlad
Singeorzan, Matthias Schmid

► **To cite this version:**

Dominik Klein, Thomas Zinner, Kathrin Borchert, Stanislav Lange, Vlad Singeorzan, et al.. Evaluation of Video Quality Monitoring Based on Pre-computed Frame Distortions. 19th Open European Summer School (EUNICE), Aug 2013, Chemnitz, Germany. pp.100-111, 10.1007/978-3-642-40552-5_10 . hal-01497007

HAL Id: hal-01497007

<https://inria.hal.science/hal-01497007v1>

Submitted on 28 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Evaluation of Video Quality Monitoring based on Pre-computed Frame Distortions

Dominik Klein¹, Thomas Zinner¹, Kathrin Borchert¹, Stanislav Lange¹, Vlad Singeorzan², and Matthias Schmid²

¹ University of Wuerzburg

² Infosim GmbH & Co. KG, 97076 Wuerzburg

Abstract. A large fraction of the current Internet traffic is caused by video streaming. Due to the growing expectations of video consumers, monitoring video applications is getting more and more important for network and service providers. In a previous work, we proposed a video quality monitoring solution which utilizes the full reference SSIM metric to improve the monitoring in the network by distributing pre-computed distortion information induced by frame losses. To improve scalability, we introduced a less complex algorithm which infers the distortion for higher loss scenarios from single loss scenarios and inter-frame dependencies. In this work, we evaluate the accuracy of our algorithm by comparing it with the exact calculation of the SSIM metric for different frame loss scenarios. We further consider different high definition test video sequences and group of picture structures and investigate the influence on the accuracy of our proposed approximation.

1 Introduction

Among the currently popular Internet applications, video streaming is responsible for the largest fraction of the global Internet traffic and is said to keep its pace within the next years [1]. This trend underlines the growing importance of video streaming in current and future networks. In parallel, technologies like software defined networking or network virtualization enable the development of application-specific virtual networks which fulfill the special requirements of applications like video streaming. Virtual networks introduce new management mechanisms [2] to optimize the virtual network towards the hosted application but require an accurate monitoring solution to assess the effects of applied management mechanisms with respect to the perceived service quality of end customers. In [3], we proposed a monitoring solution that uses a full reference metric to pre-compute the distortion per group of pictures (GOP) for different frame loss scenarios. This pre-computed information is used to improve the accuracy of the monitoring in the network, which infers the video quality from lost frames. In particular, video dependencies are included in our approach as they are captured by the video quality assessment (VQA) metric, in our case the structural similarity (SSIM) [4] metric. However, including all possible frame loss combinations per GOP introduces a large number of frame loss scenarios and hence, excessive computing power is required. To achieve a better scalability of our approach, higher frame loss scenarios are approximated by adding the

distortion of single frame loss scenarios. Hence, only the distortions for single frame loss scenarios need to be pre-computed. This approach however reduces the accuracy compared to the exact SSIM metric.

In this work, we evaluate the accuracy of our solution by comparing it with the exact calculation of the SSIM metric for different frame loss scenarios. We further consider different high definition test video sequences and GOP structures and investigate the influence on the accuracy of our proposed approximation. The remainder is structured as follows. In Section 2, we discuss similar approaches. Section 3 briefly introduces our proposed monitoring framework and in Section 4, we explain the setup of the evaluation. In Section 5, we assess the accuracy of our approach and investigate the influence of different video structures. Finally in Section 6, we conclude the paper and present future work.

2 Related Work

The quality of experience (QoE) of consumers has several influence factors like the physical and social context, the expectation and usage history of the human user, and the technical system itself [5]. Despite the technical network parameters, these factors are hard to measure. Another important factor is the video quality itself and there are several monitoring solutions which try to infer the video quality from technically measurable parameters. The most simple mechanism is to define a packet loss threshold for the IPTV service and assume the video quality as acceptable as long as the threshold is not exceeded. This technique does not take any video and content information into account. While a lost packet will produce a large error in regions with medium motion, it may produce no sizable error in regions with low motion. The mechanism introduced by Reibman et al. [6] focuses on no reference methods which estimate the video quality on network level and, if possible, on codec level. The estimation on codec level includes for instance spatio-temporal information and effects of error propagation. Tao et al. [7] propose a relative quality metric, rPSNR, which allows the estimation of the video quality against a quality benchmark provided by the network. The introduced system offers a lightweight video quality solution. Naccari et al. [8] introduce a no reference video quality monitoring solution which takes spatio-temporal error propagation as well as errors produced by spatial and temporal concealment into account. The results are mapped to SSIM and compared to results gained by computing the SSIM of the reference video and the distorted video. All these video quality monitoring mechanisms work on no reference or reduced reference metrics for estimating the video quality. A brief overview on current research questions within the area of IPTV monitoring can be found in [9].

3 Proposed Monitoring Solution

In this section, we give a brief overview to our proposed video monitoring solution. The idea of our proposed monitoring solution is to distribute pre-computed

distortion information induced by losing frames to monitoring agents in the network. The agents monitor lost frames in the video streams and utilize the pre-computed distortion information to infer the distortion of multiple frame losses within a GOP based on the distortion of single frame losses and the frame dependencies. The distortion per frame for the single frame loss scenarios and the frame dependencies are extracted on a per GOP basis during a SSIM-based video analysis prior to the video streaming process. More details to the different building blocks are provided in the following.

3.1 Precomputation of Distortion

The distortion values are computed according to the SSIM metric and we define the distortion as the dissimilarity of two frames. For each frame within a GOP, the video analysis generates a loss scenario where only this specific frame is dropped and the resulting distortion on all frames within that group is investigated. Therefore, we directly compare the undistorted image f_{Good} with the distorted image f_{Bad} via the SSIM method and hence obtain, how different the undistorted and distorted image are. The SSIM metric yields values between 0 and 1 and the distortion value per frame d_{Frame} is defined in Equation 1.

$$d_{Frame} = 1 - SSIM(f_{Good}, f_{Bad}) \quad (1)$$

The distortion value per single frame d_{Frame} hence has a maximum of 1 which means two completely different pictures. However, only I-frames are completely independent of other frames and constitute fixed pictures. All other frame types are dependent on other frames and if these frames are lost, the dependent frames cannot be decoded. Hence, a single frame can have a much higher distortion value in case a lot of other frames are dependent on this frame. To normalize the distortion per group d_{GOP} , we divide it by the number of frames per group. To get the dependencies between the frames in a GOP, we also investigate in the above emulated loss scenarios which other frames are also distorted in the currently considered GOP if a specific frame is lost.

3.2 Calculation of Video Distortion

The distortion value is calculated per GOP and once the agent sees the next GOP in the stream, the old distortion value of the former GOP is sent to the monitoring database and the value is reset to 0 for the next group. For each lost frame per group, the monitoring agent updates the distortion value d_{GOP} . First, the monitoring agent checks whether the lost frame is dependent on other frames. If the lost frame is independent, the agent looks up the distortion value for the lost frame d_{Frame} and adds this value to the distortion value of the currently considered GOP (d_{GOP}) and the update process is finished. If otherwise the lost frame is dependent on other frames, the agent needs to check whether these frames are also lost. If the currently considered frame requires another frame which is also lost, the distortion of the current frame is already included and can

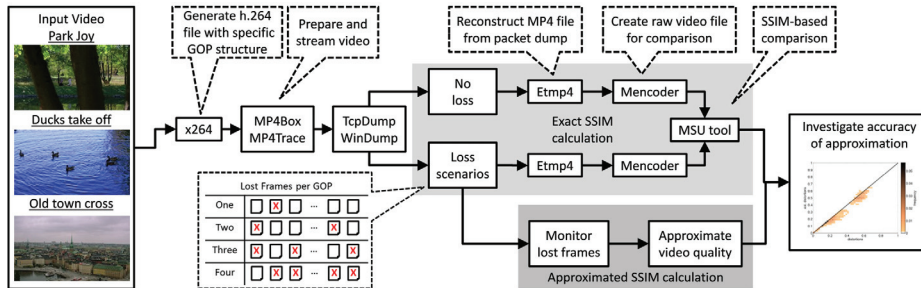


Fig. 1. Flow chart showing the different steps during the evaluation.

be ignored. In this case, no update of the d_{GOP} value is required. If in contrast the required frame is not lost, the distortion of the currently considered lost frame is not yet included and hence, the distortion d_{Frame} is added to the d_{GOP} value.

3.3 Mapping from Distortion to Video Quality

After the distortion per GOP d_{GOP} has been calculated, the monitoring needs to map the distortion to a proper metric showing the actual video quality. For our approach, the distortion is mapped to the MOS value according to [10] and then to the video quality according to [5]. There the authors have shown via subjective tests for web services that 90 % of the users already accept a fair video quality (MOS 3). A distortion per GOP $d_{GOP} \leq 0.12$ corresponds to MOS values equal or larger than MOS 3. Hence, our monitoring solution rates GOPs with a distortion $d_{GOP} \leq 0.12$ as good or accepts the video quality and rates GOPs with a distortion $d_{GOP} > 0.12$ as bad or rejects the video quality.

4 Evaluation Setup

In this section, we first describe the test video sequences and GOP structures which have been considered for the evaluation. Second, we explain the evaluation setup and the used tools.

4.1 Test Video Sequences

In the following, we introduce the considered videos for the evaluation. The videos have been selected so that different amounts of temporal and spatial information are represented. Temporal information includes the motion between consecutive frames while spatial information includes the amount of details per single frame, as introduced in ITU-T Recommendation P.910 [11]. Our evaluation was performed using free available high definition 1080p test video sequences from the Xiph.org Test Media website [12]. The videos are listed in descending

order according to the amount of temporal and spatial information and correspond to different content types such as high or low motion and high or low detail. The considered sequences are *Park Joy*, *Ducks Take Off*, and *Old Town Cross*. Screenshots for the three different videos are shown on the left-hand side in Figure 1.

4.2 Assessed GOP Structures

In addition to the influence of different types of videos, we also investigate how well our proposed monitoring solution behaves for different GOP structures. Therefore, we have analyzed the GOP structures currently used for live streaming of video content by two prominent German IPTV broadcasters, i.e. the German Telecom and the German public service broadcasters (ARD/ZDF). The German Telecom offers an IPTV service called T-Entertain which can be booked in addition to the DSL Internet connection. Entertain runs in a separate VLAN and provides access to various TV channels in high and standard definition quality. The GOP structure for Entertain in HD quality is $M = 8, N = 64$, whereas M denotes the distance between P-frames and N the distance between I-frames. Hence, for Entertain HD, there are 7 B-frames after an I- or P-frame and between two I-frames, there are 63 P- and B-frames. This kind of GOP structure reduces the amount of transmitted information but the encoding and playout order of frames is different which increases the complexity at decoder side. For the ARD/ZDF live TV stream in contrast, the GOP structure is different. There are no B-frames at all and also the length of the GOP is variable and not fixed. A possible explanation for the variable GOP length is that the length is adaptive to the video content to reduce the video bit rate. Such an approach is for example proposed in [13]. Hence, the GOP structure for the ARD/ZDF live stream can be written as $M = 1, N = \text{variable}$. This kind of structure has an increased video bitrate compared to the structure with B-frames but the encoding and playout order of frames is the same.

Table 1. Considered GOP structures.

Label	Size	Length	Structure
IBP	M=4	N=16	IBBBPBBBBPBBBBPBBB
IPP	M=1	N=16	IPPPPPPPPPPPPPPPPP

For our evaluation, we use slightly adapted GOP structures considering the high definition video test sequences. The GOP length for both structures has been set to 16 frames and is not dependent on the video content anymore so that both structures are comparable. For the structure with B-frames, we have reduced the GOP size to 4 which results in 3 B-frames in between I- and P-frames. The resulting modified GOP structures can be seen in Table 1. For convenience, the structure with B-frames is denoted as *IBP* structure in the following and the structure without B-frames is denoted as *IPP* structure.

4.3 Evaluation Methods

For the evaluation of the accuracy of our proposed monitoring solution, we compare our approximated SSIM values with the exact SSIM values for different frame loss scenarios. The overall setup as well as the different steps during the evaluation can be seen in Figure 1. In the first step, the x264 tool [14] is used to create a h.264 file with a specific GOP structure (see Table 1) for the three different input videos. In the second step, we use the MP4Box tool [15] and the MP4Trace tool from the EvalVid framework [16] to create a video stream which is dumped in the next step using either TcpDump [17] or WinDump [18]. To evaluate different frame loss scenarios, we create a lossy dump file in the next step by removing certain frames from the dump file. This way, we generate different dump files for loss scenarios where exactly one, two, three, or four frames are lost per GOP. This results in $\binom{16}{i}$ scenarios in case exactly i frames are lost per GOP. Evaluating all frame loss combinations results in a high number of scenarios which requires excessive computing power. Hence, we limit our evaluation to at most four frame loss scenarios and show that higher loss scenarios are not required to demonstrate the accuracy of our approach. After the different dump files have been created, we use the Etmp4 tool from the EvalVid framework [16] to reconstruct the MP4 file from the dump file and create raw video files by using the MEncoder tool [19]. In the last step, we use the MSU tool [20] to compare the lossy video file with the original video file by calculating the exact SSIM values. These values are then used as reference and we evaluate the induced error due to our approximation of the SSIM metric. Therefore, the information about the lost frames is used by our proposed monitoring solution to calculate the approximated SSIM values (see lower part of Figure 1). The approximated SSIM values are then compared with the exact SSIM values calculated by the MSU tool.

5 Evaluation of Video Quality Monitoring

In this section, we show the accuracy of our approach with respect to the exact SSIM metric. First, we investigate which frame loss scenarios are relevant and which higher scenarios cannot deliver a good video quality anymore. Second, we assess the influence of the different video types and GOP structures on our approach and show, how our metric can be optimized for certain GOP structures. Third, we further investigate the error due to our approximation and finally, we show the sensitivity of our approach with respect to the acceptance threshold for the video quality.

5.1 Relevant Frame Loss Scenarios

In this subsection, we investigate which loss scenarios lead to a large fraction of GOPs where the video quality is still acceptable. For these scenarios, our proposed monitoring solution needs to be accurate. For loss scenarios where

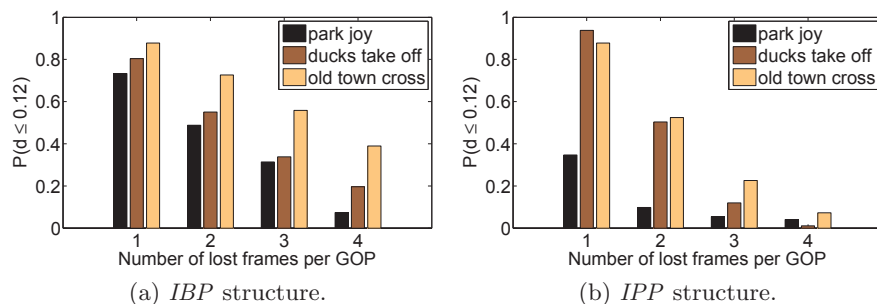


Fig. 2. Number of GOPs with distortion value less than 0.12.

a large fraction of GOPs has a high distortion and hence a very bad quality, accuracy is not that important. According to Section 3.3, our mapping rates the video quality of a GOP acceptable if the distortion per GOP $d_{GOP} \leq 0.12$.

The resulting percentage of GOPs with a distortion less than 0.12 for the three different videos is shown in Figure 2. Figure 2a shows the results for the *IBP* structure and Figure 2b shows the corresponding results for the *IPP* structure. In both figures, the x-axis shows the number of lost frames per GOP and the different colored bars denote the three different videos.

Concerning the *IBP* structure in Figure 2a, we see that the fraction of GOPs with acceptable quality decreases for the higher loss scenarios. This is in line with the expectations because the more frames are lost, the worse is the overall video quality per GOP. However, there are strong differences between the different types of video which can be explained due to the amount of spatial and temporal information. The Park Joy video sequence has the highest amount of information and is hence more susceptible to frame loss than the other two videos which have a higher number of acceptable GOPs in all loss scenarios. Even if four frames are lost within a GOP, about 40 % of the GOPs for the Old Town Cross test sequence still have an acceptable video quality. However, for five and six frame loss scenarios, the percentage of GOPs with good quality drops to 24.3 % and 13.32 % respectively. Hence, it would be necessary to take five and six frame loss scenarios into account. However, due to the high computational complexity we omit these scenarios for the video clip Old Town Cross.

For the *IPP* structure depicted in Figure 2b, similar observations are made. However, the fraction of acceptable GOPs is much smaller for all videos in the higher loss scenarios. Due to the absence of B-frames, the overall importance per frame is higher and hence, this structure is more susceptible to frame loss than the *IBP* structure. If we directly compare the number of accepted GOPs for the four frame loss scenario, we see that the *IPP* structure in the right figure has a much lower fraction as the *IBP* structure on the left figure. Overall, also for the *IPP* structure, it is sufficient to consider only loss scenarios where at most four frames are lost per GOP. For higher loss scenarios, our approach indicates an unacceptable video quality with a high probability.

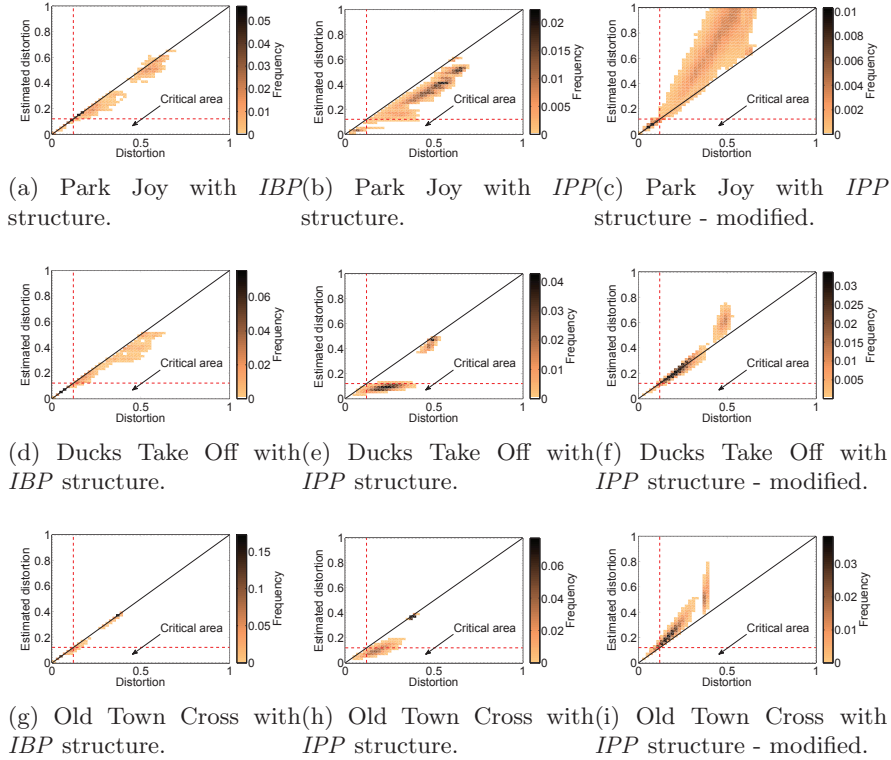


Fig. 3. Scatter plots showing estimated distortion against exact distortion values.

5.2 Qualitative Evaluation of Accuracy

In this subsection, we plot our approximated distortion values against the distortion values calculated by the exact SSIM metric to show the accuracy of our monitoring solution. The evaluation does only consider loss scenarios where exactly two, three, and four frames are lost. One frame loss scenarios are not considered as for those scenarios, our approximation yields the same results as the exact SSIM metric. The results for the different video test sequences and GOP structures can be seen in Figure 3. Figures 3a-3c show the results for the Park Joy sequence, Figures 3d-3f show the results for the Ducks Take Off sequence, and 3g-3i show the results for the Old Town Cross sequence. All figures are plotted as scatter plots where the x-axis denotes the exact distortion values and the y-axis denotes the estimated distortion values. All plots also contain the 0.12 threshold lines for the estimated and exact distortion values as well as the identity line through the origin. Estimated distortion values which lie on the line through the origin perfectly match the exact distortion values.

Considering the *IBP* structure (see Figures 3a, 3d, and 3g) for the three different videos, we see that our approach performs best for the Old Town Cross video which has the lowest spatial and temporal information. For the other two

video sequences, our approximation still performs very well in the critical area around the video quality acceptance threshold of 0.12 and only deviates in the higher distortion areas. There however, our monitoring does not accept GOPs with a bad video quality or a distortion value higher than the acceptance threshold and hence, no error occurs. A different observation can be seen for the *IPP* structure (see Figures 3b, 3e, and 3h). For this structure, our proposed approximation does not perform well and underestimates the distortion in the critical area around the acceptance threshold for all three videos. For this structure, all subsequent frames are always dependent on their precedent frames and errors in earlier frames influence all subsequent frames. Our approximation however ignores the distortion values for frames which are dependent on earlier frames (see Section 3.2) as the distortion of dependent frames is included in the distortion value of their required frame. This is a good approximation for GOP structures with minor inter-frame dependencies like the *IBP* structure but not for *IPP* structure. Hence, to improve our approach, we modify the calculation of the distortion per GOP d_{GOP} and always add the distortion of lost frames d_{Frame} instead of ignoring the distortion of frames which are dependent on another lost frame. This is a very simple modification to our initial metric. But for the specific use case, the results prove the viability of this approach.

Figures 3c, 3f, and 3i show the results for the modified version of our approximation. With the modified approach, we do not underestimate the distortion in the critical area. Accordingly, a large fraction of the estimations lies on the identity line through the origin. For the higher distortion area, our modified approach now overestimates the distortion. There however, our monitoring does not reject GOPs with a good video quality or an exact distortion smaller than 0.12 and hence, no error occurs.

5.3 Quantitative Evaluation of Accuracy

The scatter plots in the former section give a basic understanding about how our unmodified and modified metric perform for the different videos and GOP structures. However, for a quantitative statement, CDF plots are more suitable. Hence in Figure 4 and 5, we plot the CDF for the error e between the estimated and the exact distortion values. The error e is defined as the exact distortion minus the estimated distortion. A negative error means that our proposed approximation overestimates the distortion and positive error means an underestimation of distortion. From the perspective of a network provider, a positive error is more serious because the monitoring underestimates the distortion and hence recognizes a bad video quality too late or even not at all.

Regarding the *IBP* structure, we again see that our proposed monitoring solution performs best for the Old Town Cross sequence. For that video, all GOPs have an error $e < 0.05$. For the other two videos, about 80 % of the GOPs have an error $e < 0.05$. However, as we have seen in Figure 3, the larger error occurs in a distortion range where accuracy is less important.

For the *IPP* structure, it can be seen that the unmodified approach always significantly underestimates the distortion and hence is not suitable for this

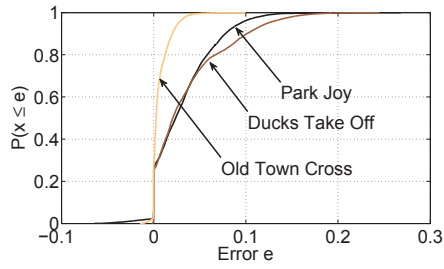


Fig. 4. Error depicted as CDF for *IBP* structure.

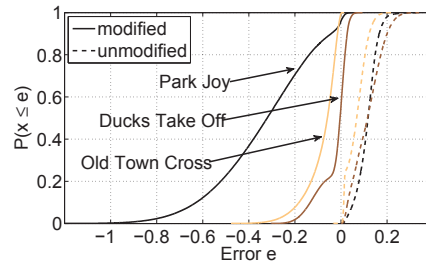


Fig. 5. Error depicted as CDF for *IPP* structure.

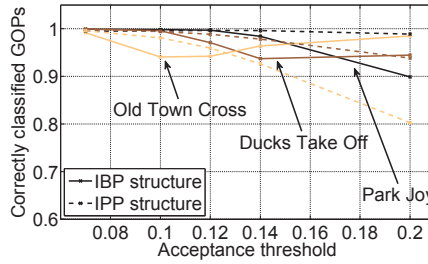


Fig. 6. Sensitivity with respect to acceptance threshold.

structure. In contrast, the modified approach mostly overestimates the distortion and only a very small fraction of GOPs has a positive error e . The modified approach is hence more suitable for the *IPP* structure as the GOPs with a high negative error e have a very high distortion and are rejected anyway.

5.4 Sensitivity with Respect to Acceptance Threshold

In the former evaluation of the accuracy of our proposed monitoring solution, we have used an acceptance threshold of 0.12 for the video quality. The video quality of GOPs is only accepted if the corresponding distortion is less than this threshold. However, this threshold was chosen according to [5], where the authors have shown via subjective tests for web services that 90 % of the users accept a fair service quality (MOS 3). If due to new findings in future work this threshold needs to be adapted, our monitoring solution should still be accurate. Hence in the following, we investigate how the percentage of correctly classified GOPs behaves for different acceptance thresholds. The corresponding results are shown in Figure 6. The x-axis shows different values for the acceptance threshold and the y-axis shows the percentage of correctly classified GOPs. The solid lines denote the results for the *IBP* structure and the dashed lines denote the results for the *IPP* structure. For the *IPP* structure, only the results with the modified approach are shown.

For all structures and videos, a threshold close to 0 leads to 100 % correctly classified GOPs. In that case, both metrics always reject the video quality for a GOP if the distortion is larger than 0 which is not a reasonable approach as such a monitoring would be far too pessimistic. For the chosen threshold of 0.12, our monitoring classifies about 98 % of the GOPs correctly. Only the *IBP* structure for the Old Town Cross sequence experiences a slightly lower classification rate of 93 %. For an increasing threshold, the classification rate for nearly all videos and structures decreases. For the *IBP* structure, the classification rate does not drop below 90 % which is still an acceptable result. For the *IPP* structure, the classification rate drops to 80 % for the Old Town Cross video. Overall, our monitoring still achieves a high correct classification of GOPs even in the higher distortion range of about 0.2.

6 Conclusion

In this work, we presented an evaluation of our proposed monitoring solution for IP video streaming services which utilizes knowledge about the video content to predict the service quality. Our solution pre-computes the distortion induced by losing frames using the full reference metric SSIM. Due to the high complexity for computing all frame loss possibilities within a Group of Pictures (GOP), we introduce a less complex algorithm which computes the distortion of multiple frame losses within a GOP based on the distortion of single frame losses and the frame dependencies. We investigate the accuracy of the introduced approximated monitoring solution for two different video streaming configurations used by German broadcasting services and the Deutsche Telekom.

Our results indicate an accuracy of more than 95% correctly classified GOPs of the proposed approximative distortion computations as compared to the correct values. At the same time the number of required computation is significantly reduced, since only single frame losses within a GOP have to be computed. Future work will focus on the comparison of the proposed solution with other monitoring approaches from literature with respect to the trade-off between monitoring costs, scalability, and accuracy. This also includes subjective user surveys which can be used as a metric for the comparison, and also to improve the accuracy of the proposed video monitoring solution.

Acknowledgment

The authors would like to thank Prof. Tran-Gia for the fruitful discussion and the support in this work.

References

1. C. S. Inc., “Cisco visual networking index: Forecast and methodology, 2011-2016,” http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf, June 2012.

2. S. Meier, M. Barisch, A. Kirstädter, D. Schlosser, M. Duelli, M. Jarschel, T. Hoßfeld, K. Hoffmann, M. Hoffmann, W. Kellerer, A. Khan, D. Jurca, and K. Koza, "Provisioning and Operation of Virtual Networks," *Electronic Communications of the EASST, Kommunikation in Verteilten Systemen 2011*, vol. 37, Mar 2011.
3. D. Klein, T. Zinner, S. Lange, V. Singeorzan, and M. Schmid, "Video Quality Monitoring based on Precomputed Frame Distortions," in *IFIP/IEEE International Workshop on Quality of Experience Centric Management (QCMAN)*, Ghent, Belgium, May 2013.
4. Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.
5. R. Schatz, T. Hoßfeld, L. Janowski, and S. Egger, "From Packets to People: Quality of Experience as New Measurement Challenge," in *Data Traffic Monitoring and Analysis: From measurement, classification and anomaly detection to Quality of experience*, M. M. Ernst Biersack, Christian Callegari, Ed. Springer Computer Communications and Networks series, 2012.
6. A. Reibman, V. Vaishampayan, and Y. Sermadevi, "Quality monitoring of video over a packet network," *IEEE Transactions on Multimedia*, vol. 6, no. 2, Apr 2004.
7. S. Tao, J. Apostolopoulos, and R. Guerin, "Real-Time Monitoring of Video Quality in IP Networks," *IEEE Transactions on Networking*, vol. 16, no. 6, Dec 2008.
8. M. Naccari, M. Tagliasacchi, and S. Tubaro, "No-Reference Video Quality Monitoring for H.264/AVC Coded Video," *IEEE Transactions on Multimedia*, vol. 11, no. 5, Aug 2009.
9. J. Apostolopoulos and A. Reibman, "The Challenge of Estimating Video Quality in Video Communication Applications [In the Spotlight]," *Signal Processing Magazine, IEEE*, vol. 29, no. 2, pp. 160–158, 2012.
10. Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment using structural distortion measurement," in *International Conference on Image Processing*, vol. 3, 2002, pp. 65–68.
11. I. T. Union, "ITU-T Recommendation P.910: Subjective video quality assessment-methods for multimedia applications," <http://www.itu.int/rec/T-REC-P.910/en>, Apr 2008.
12. xiph.org, "Derf's test media collection," <http://media.xiph.org/video/derf/>, Mar 2013.
13. B. Zatt, M. Porto, J. Scharcanski, and S. Bampi, "Gop structure adaptive to the video content for efficient H.264/AVC encoding," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, Sep 2010, pp. 3053–3056.
14. "x264 - h264/avc encoder," <http://www.videolan.org/developers/x264.html>.
15. G. multimedia open source project, "Mp4box," <http://gpac.wp.mines-telecom.fr/mp4box/>.
16. J. Klaue, B. Rathke, and A. Wolisz, "EvalVid - A Framework for Video Transmission and Quality Evaluation," in *Proc. of 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, Urbana, Illinois, USA, Sep 2003.
17. "tcpdump/libcap," <http://www.tcpdump.org/>.
18. "windump/winpcap," <http://www.winpcap.org/windump/>.
19. T. M. Project, "Mencoder," <http://mplayerhq.hu/design7/news.html>.
20. Graphics and M. S. U. Media Lab, CMC department, "Msu video quality measurement tool."