SSIM-based video admission control and resource allocation algorithms

Marco Zanforlin†, Daniele Munaretto†, Andrea Zanella†, Michele Zorzi†*
† Department of Information Engineering, University of Padova, Italy
* Consorzio Ferrara Ricerche, Ferrara, Italy
E-mail: {zanforli, munaretto, zanella, zorzi}@dei.unipd.it

***Abstract*—The exponential growth of video traffic in mobile networks calls for the deployment of advanced video admission control (VAC) and resource management (RM) techniques in order to provide the best quality of experience (QoE) to the end user according to the available network resources. The degradation of the QoE perceived by the user when reducing the source rate of a video typically depends on the content of the video itself. In this paper, we analyzed the QoE of a group of test video sequences encoded with H.264 advanced video codec at different rates, i.e., quality levels. The QoE is objectively expressed in terms of the average structural similarity (SSIM) index. Based on empirical results, we propose a 4-degree polynomial approximation of the SSIM as a function of the coded video rate. We hence propose to tag each video with these polynomial coefficients that provide a compact description of its specific SSIM behavior, and to use this information in VAC and RM algorithms to optimally manage a shared transmission medium. As a proof of concept, we report selected simulation results that compare QoE-aware and QoE-agnostic algorithms in a scenario with a single link shared by multiple concurrent video flows.**

***Index Terms*—QoE; video delivery; SSIM; resource management; video admission control.**

I. Introduction

Mobile data and video services are quickly becoming an essential part of consumers lives. In 2012 the mobile video traffic already exceeded 50% of the total data traffic in the Internet. The latest global mobile data traffic forecast from [1] foresees a further increment of the video traffic growth of 75% between 2012 and 2017, accounting for over 66% of the total mobile data traffic by the end of the forecast period. Because mobile video contents are highly bit-rate demanding they will generate most of the cellular traffic growth in the near future. Backhaul capacity must increase so that mobile broadband, data access, and video services can effectively support consumer usage trends and keep mobile infrastructure costs at a reasonable level.

At the current stage, mobile network operators face the issue of supporting high quality video services with the available network resources. The widespread high-speed wireless coverage, e.g., by means of femto-cells and WiFi hotspots, will likely increase the number of users that require high quality *mobile* video services, making the support of such a traffic in the access networks a challenging question.

An attractive solution in this scenario consists in dynamically adapting the QoE perceived by the final video consumers to the available transmission resources by adjusting the video code rates. As observed in [2], in fact, reducing the encoding rate of a video is much less critical in terms of QoE degradation than increasing the packet loss probability or the delivery delay. However, the perceived QoE at a certain encoding rate depends on the specific characteristics of the video, e.g., scene and source dynamics and frame-by-frame motion.

In this paper, we propose a novel approach to handle under-provisioned video traffic scenarios that is based on the possibility of dynamically adjusting the video encoding rate and that takes into account the different QoE behaviors of video sequences, expressed in terms of, e.g., SSIM index [3]. QoE-based full-reference metrics such as Video Quality Model (VQM) [4] and MOtion-based Video Integrity Evaluation (MOVIE) index [5] suit well for catching the temporal consistency of the video frames. Nevertheless, for the sake of simplicity, we adopt SSIM as human visual system-similarity measure to assess the video quality, since it is widely adopted for video quality testing in the community [6]–[8]. We stress the point that the framework we design in this paper is independent of the QoE-based metric in use, thus it can be straightforwardly adapted to metrics which take into account specific video features of interest.

We measure the SSIM of a large set of H.264-AVC [9] video clips coded at different rates, which correspond to different perceived quality levels. After a suitable normalization and rescaling of the metrics of interest, we are able to analytically approximate the perceived quality characteristic of each video by means of a simple 4-degree polynomial expression. We hence associate each video to its polynomial coefficients in order to compactly describe how the SSIM degrades for lower transmit rates. To reduce the complexity of the video tagging, we also proposed a class-based approach where videos showing similar SSIM *vs* rate relations, inspected via quasi-real time investigation methods such as, e.g., neuronal networks, are grouped in a class and tagged with a set of polynomial coefficients that characterize that class. We then define RM and Video Access Control VAC algorithms that take into account this QoE-based information associated to each video to maximize a certain utility function. As a proof of concept, we apply our approach to a simple network scenario with a congested link shared by multiple video streams. By means of simulation, we show promising results for realistic deployments of video streaming services over wireless and cellular networks.

The remainder of the paper is organized as follows. In Section II we review the related work. Our analysis is presented

978-3-901882-63-0/14/$31.00 ©2014 IEEE

in Section III, followed by the design of our algorithms in Section IV. The simulation results are reported in Section V and we conclude the paper in Section VI.

II. Related Work

Prior works on video classification mainly focus on extracting objective networking and quality metrics. In [10] the authors classify videos based on selected common spatial-temporal audio and visual features described by the MPEG-7 compliant content descriptors. Due to the complexity of the method, the authors make use of the principal component analysis to reduce the set of features. Nevertheless, this work is strictly dependent on the MPEG-7 format. Scene detection mechanisms were developed based on predictive analytical models as in [11]. The authors propose a scene-change detector for video-conference traces that works based on the average number of bits generated during the scenes, and it is modeled with a two-state Markov chain. The proposed low complexity method comes at the cost of requiring full knowledge of the type of video to properly set the thresholds for the scene recognition. Further related works focus on quality prediction models to capture the behavior of video scenes. In [12], an objective model to predict the quality of the lost frames for 3D videos is designed based on the header information of the video packets at different ISO/OSI layers. This model roughly captures the SSIM of some video clips based on the size of the lost frames and via deep packet inspection, which is usually avoided by operators due to the complexity and national privacy rules. Nevertheless, in [13], the authors claim that the frame loss probability provides only a limited insight into the video quality perceived by the user. Moreover, the authors state that the rate distortion curves drawn using the Peak Signal-to-Noise Ratio (PSNR) provide a limited representation of the perceived video quality, thus improved quality metrics to better represents videos are needed.

In our work, we group video sequences based on the relation between video compression rate and SSIM. As widely recognized, SSIM improves traditional objective QoS metrics like PSNR and mean square error (MSE), which have been proven to be inconsistent with the human eye perception. SSIM is widely used to assess the video quality [6]–[8], specially when the scope of the work is to express the perceived quality of sequential static images coded at different levels of compression, as in our work. However, it does not take into account the impact of the temporal consistency of video frames on the perceived quality [14], which is out of the scope of this paper. We show that the SSIM characterization of a video sequence can be compactly represented by means of only four polynomial coefficients, which can be associated to the video. Tagged videos can then be handled by simple traffic shaping mechanisms in case of network congestion or under-provisioned network resources. Furthermore, to reduce the complexity of the RM and VAC unit, we propose to cluster the videos into a small number of QoE-based classes, at the cost of a marginal performance degradation. We observe that the brute-force computation of the SSIM coefficients for each video can actually be computationally expensive since it requires to evaluate the SSIM index for the entire video for the different encoding schemes. However, we speculate that machine learning techniques can make it possible to estimate these coefficients based on macroscopic characteristics of the video clip, with much lower computational complexity, though this approach is still under investigation.

TABLE I
Mapping SSIM to Mean Opinion Score scale

SSIM	MOS	Quality	Impairment
≥ 0.99	5	Excellent	Imperceptible
$[0.95, 0.99)$	4	Good	Perceptible but not annoying
$[0.88, 0.95)$	3	Fair	Slightly annoying
$[0.5, 0.88)$	2	Poor	Annoying
< 0.5	1	Bad	Very annoying

III. Empirical analysis of videos

A. Setup

We evaluate the QoE of the videos with the SSIM index. From the human eye perspective, SSIM improves the representation of the perceived video quality compared to traditional metrics such as PSNR and MSE. These pure mathematical metrics estimate the perceived distortion based on analytical pixel-by-pixel comparison. On the contrary, SSIM measures the image degradation in terms of perceived structural information change, thus taking into account the tight interdependence between spatially close pixels which contain the information about the objects in the visual scene. SSIM is calculated via statistical metrics (mean, variance) computed within a square window of size $N \times N$ (typically 8×8), which moves pixel-by-pixel over the entire image. The measure between the corresponding windows X and Y of two images is computed as follows:

$$SSIM(X,Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \quad (1)$$

with μ and σ^2 the mean and variance of the luminance value in the corresponding window, and c_1 and c_2 variables to stabilize the division with weak denominator (we refer the interested reader to [3] for more details on the computation). For practical reasons, we take the average values of SSIM for each video. The range of the SSIM index goes from 0 to 1, which represent the extreme cases of totally different or perfectly identical frames, respectively. Tab. I shows the mapping between SSIM and Mean Opinion Score (MOS) scale, as reported in [15]. We remark that SSIM captures the *spatial* differences between two representations of the *same frame* and, hence, is particularly suitable to express the perceived quality of a *static* image when coded at different levels of compression. This metric does not consider the effect of the temporal correlation between consecutive frames in a video clip, which may alleviate the perceptual impact of compression artifacts in the video frames. Nonetheless, many studies have shown that the *average* SSIM computed over a sequence of frames of a video clip is generally a good QoE index for the video as well [6]–[8]. This is also confirmed by the result of our study, as discussed next.

B. Video evaluation

We consider a pool of $V = 38$ CIF video clips, taken from standard reference sets.[1] Each video has been encoded with the

[1]Video traces can be found in [16], ftp://132.163.67.115/MM/cif

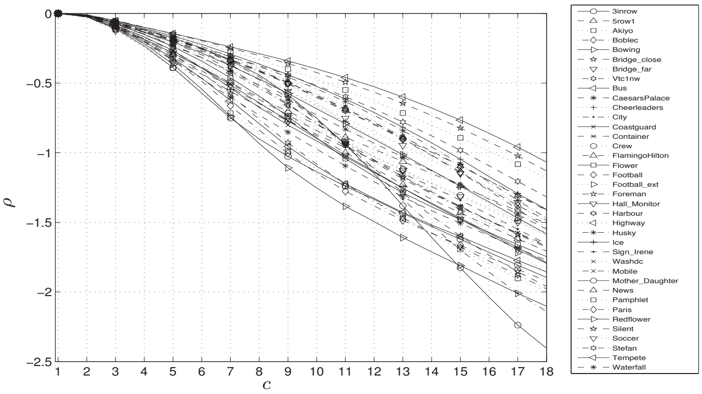


Fig. 1. Logarithm of the normalized rate $\rho_v(c)$ *versus* compression level c for different video clips.

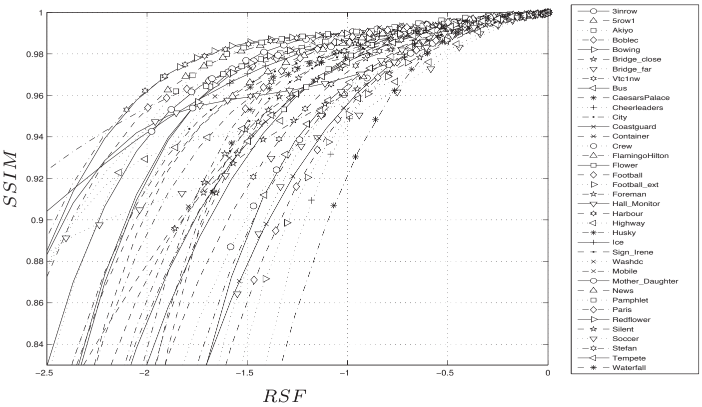


Fig. 2. SSIM of the different video clips when varying the RSF.

Joint Scalable Video Model (JSVM) reference software [17] into H.264-AVC format at $C = 18$ increasing compression rates, which correspond to as many quality levels. The list of video names, full quality transmit rate, duration and classification of video motion is provided in Tab. II at the end of the paper. We denote by $c \in \{1, \ldots, C\}$ the available compression rates and by $r_v(c)$ the transmit rate of video $v \in \{1, \ldots, V\}$ encoded at rate c, with $r_v(1)$ being the maximum (i.e., full quality) rate. To ease the comparison between different video clips, it is convenient to normalize the video rates to the full quality rates. Moreover, following the Weber-Fechner's law that speculates a logarithmic relation between the intensity and the subjective perception of a stimulus, we introduce a logarithmic measure of the normalized rate, here named *Rate Scaling Factor* (RSF), which is defined as

$$\rho_v(c) = \log(r_v(c)/r_v(1)) , \tag{2}$$

and shown in Fig. 1 when varying the compression level c for the different videos.

We can see that the compression level, i.e., the number of quantization points considered in the H.264-AVC encoding, determines the rate of the video sequence depending on the content of the video itself. For a given compression level c, the larger the RSF $\rho_v(c)$ the more dynamic the video sequence. Indeed, dynamic sequences exhibit lower spatial and temporal correlation of consecutive video frames and, hence, are less amenable to compression. The dynamics of the video content also impact the perceived QoE for a certain RSF value, as clearly shown in Fig. 2 that shows the average SSIM of each video clip when varying ρ_v. At visual inspection, indeed, we first observe that the perceived degradation of video quality when reducing the source rate is generally coherent with the decrease rate of the SSIM, which confirms that the SSIM is a good performance index for the QoE of video sequences. Second, we notice that, reducing the source rate, the more dynamic the video, the lower the perceived video quality. In Fig. 2, markers correspond to empirical SSIM values for the V videos in the test set, while the lines are obtained from a 4-degree polynomial approximation of such values. We observe that the polynomial approximation is, in general, acceptably accurate for the range of ρ of practical interest. Therefore, the SSIM characteristic of a video v can be approximated as

$$F_v(\rho) \simeq 1 + a_{v,1}\rho + a_{v,2}\rho^2 + a_{v,3}\rho^3 + a_{v,4}\rho^4 . \tag{3}$$

The vector of coefficients $\mathbf{a}_v = \{a_{v,i}\}$ provides a compact description of the relation between the perceived QoE and the RSF of a video v. It is hence conceivable to tag each video or, when optimizing at a finer granularity, each single Group of Picture (GoP), with such a compact representation of its QoE characteristic, which can then be used by RM and VAC algorithms as discussed in the next section.

IV. SSIM-based RM and VAC Algorithms

In this section we investigate how the QoE characterization of a video sequence can be used to optimally allocate transmission resources to different video sessions and to decide whether or not a new video shall be admitted into the system.

We consider a framework where different video clips are multiplexed into a shared link of capacity R by a control unit that performs RM and VAC. More specifically, the RM module detects changes of the link capacity (e.g., due to concurrent data flows or fading phenomena in wireless channels) and triggers an optimization procedure that adapts the video rates to maximize a certain QoE-related utility function. Similarly, the VAC module determines whether or not a new video request can be accepted without decreasing the QoE of any videos below a threshold F^* negotiated, for instance, between operator and video consumers. To this end, the VAC invokes the RM module to get the best resource allocation policy for all the videos potentially admitted into the system and, then, computes the SSIM of each video through (3) and checks whether it is above the aforementioned quality threshold. If not, the last video admission request is refused, otherwise the new video is accepted into the system, transmission resources are reallocated as determined by the RM module, and video sources are required to adapt their source rate to such a new allocation. To alleviate the RM from the burden of computational costs when dealing with a high number of active videos in the channel and to make the inspection of video features feasible in practice for quasi-real time applications, we further consider a clustering approach where videos are grouped based on their SSIM vs. rate similarity. Hence, we partition the videos in K classes, depending on the value of ρ for which the SSIM crosses the threshold F^*. Each class is then associated to a reference F curve, expressed as in (3), with polynomial coefficients equal to the mean of the coefficients of the videos in that class:

$$F_k^{\mathcal{C}}(\rho) = 1 + \sum_{i=1}^{4} a_{k,i}^{\mathcal{C}} \rho^k , \quad \text{with} \quad a_{k,i}^{\mathcal{C}} = \frac{\sum_{v \in \mathcal{C}_k} a_{v,i}}{||\mathcal{C}_k||} ,$$

where $\mathcal{C}_k$ denotes the set of videos in the kth class, and $||\mathcal{C}_k||$ is the cardinality of the class. Class-based QoE functions $F^{\mathcal{C}}$ can be used by the RM and VAC algorithms in place of the actual F of each video, thus reducing the computational complexity at the cost of suboptimal resource allocation and possible violation of the SSIM constraints, due to the coarse approximation of the QoE characteristic of the videos.

A. Optimal resource allocation problem

Let f_i denote the SSIM function associated to a video i, which can correspond to either $F_i^{\mathcal{C}}$ or F_i, depending on whether or not the class-based solution is considered. Furthermore, let R denote the transmission capacity that needs to be allotted to the videos, and by $\Gamma = \{\gamma_i\}$ an allocation vector that assigns to the ith video a fraction γ_i of R, with $\gamma_i = 0$ indicating that the video is not accepted into the system. Although the H.264 encoding can only offer a discrete set of transmit rates (see Fig. 1), in the formulation of the optimization problem we assume that video rates can change in a continuous manner. Under this assumption, the RSF of the ith video can be expressed as

$$\tilde{\rho}_i = \log\left(\frac{\gamma_i R}{r_i(1)}\right) . \tag{4}$$

The optimization problem addressed by the RM module can then be defined as follows:

$$\Gamma_{\text{opt}} = \arg\max_{\Gamma} U(\Gamma, R, \{f_i\}) \quad \text{s.t.} \quad \sum_i \gamma_i \le 1 \tag{5}$$

where $U(\cdot)$ denotes the *utility function* considered by the optimization algorithm.

We consider two utility functions that reflect different optimization purposes:

Rate Fairness (RF): Resources are distributed to all active videos proportionally to their full quality rate, without considering the impact on the perceived QoE. In this case, the optimal rate allocation for the ith video is simply given by

$$\gamma_{\text{opt},i} = \frac{r_i(1)}{\sum_j r_j(1)} \tag{6}$$

so that the RSF of each video equals $\tilde{\rho} = \log(R/\sum_j r_j(1))$.

SSIM Fairness (SF): Resources are allocated according to a max-min fairness criterion with respect to the SSIM of the different videos:

$$U(\Gamma, R, \{f_i\}) = \min_i f_i(\rho_i) . \tag{7}$$

Note that, under the assumption of continuous rate adaptation, the SF criterion yields the same SSIM, say σ, to all active videos. Given this target SSIM, the RSF for each video can be easily found as $\tilde{\rho}_i = f_i^{-1}(\sigma)$, where f_i^{-1} is the inverse of the QoE monotonic function f_i. Therefore, the optimization problem can be easily solved by searching for the maximum σ that satisfies the rate constraint in (5), i.e., such that

$$\frac{1}{R} \sum_i r_i(1) 10^{f_i^{-1}(\sigma)} \le 1 . \tag{8}$$

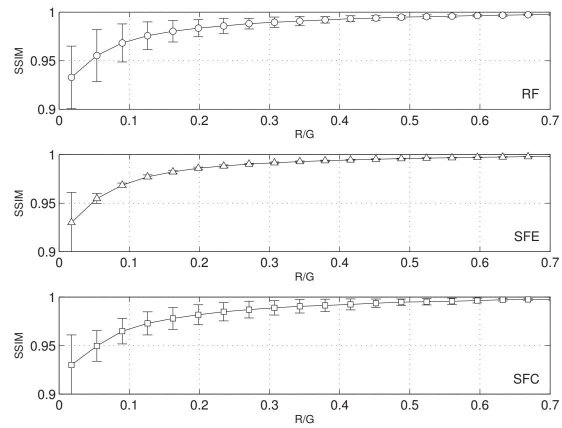


Fig. 3. Mean and standard deviation of videos' SSIM when varying the normalized channel rate R/G, for different RM algorithms.

Since the left-hand side expression is monotonic, the maximum σ can be easily and quickly found via numerical method (e.g., binary search). The optimal resource allocation is then given by

$$\gamma_{opt,i} = \frac{r_i(1)}{R} 10^{f_i^{-1}(\sigma)} .$$

B. RM and VAC algorithms

Based on these utilization functions, we can define three possible RM algorithms, namely

- `RF`, based on (6);
- `SFE` based on (7) with exact (E) QoE characterization. i.e., $f_i = F_i$;
- `SFC` based on (7) with class-based (C) QoE characterization., i.e., $f_i = F_i^{\mathcal{C}}$.

Given the channel capacity R and the set of videos to be allocated, each tagged with the polynomial coefficients associated to $\{f_i\}$ and the available encoding rates $\{r_i(c)\}$, the RM algorithm finds the optimal allocation Γ_{opt} under the continuous rate assumption and, then, looks for a feasible rate allocation at minimum distance from Γ_{opt} that satisfies (5).

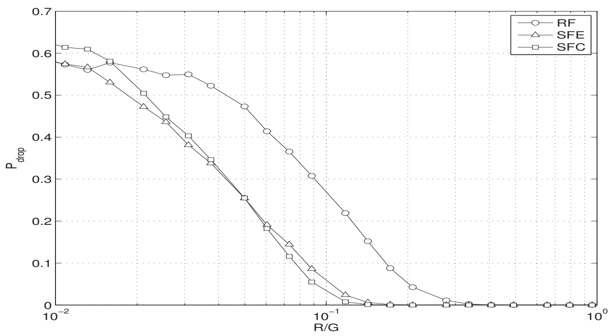
As mentioned before, the VAC algorithm can be built upon any RM since it simply calls the RM to perform its computations on the set of active videos.

V. Simulation results

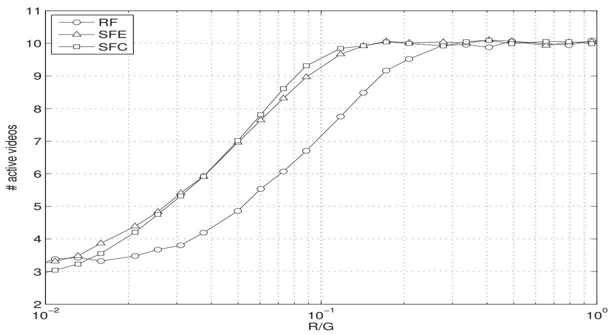
We implement in Matlab a video delivery framework in which a shared transmission channel with capacity R is controlled by an access unit that provides i) resource allocation to the different video sources and ii) video admission control functionalities.

To begin with, we compare the performance of the RM algorithms in Section IV by assuming that all $V = 38$ videos in our test set are simultaneously active when varying the capacity R of the channel. Fig. 3 reports the average SSIM when varying R for the three RM schemes. Note that the channel capacity has been normalized to the maximum aggregate traffic rate generated by the video sources, given by

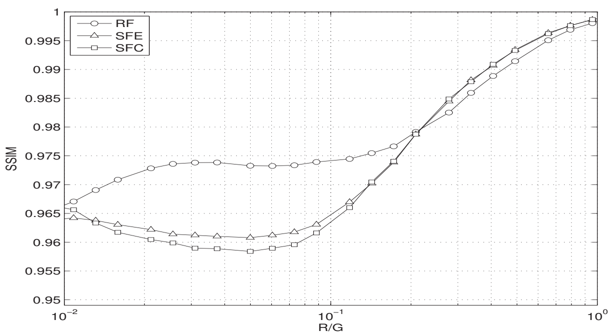
$$G = \sum_{v=1}^{V} r_v(1) . \tag{9}$$



(a) Video dropping probability *vs* channel rate.



(b) Average number of active videos *vs* channel rate.



(c) Average SSIM of active videos *vs* channel rate.

Fig. 4. VAC performance with different RM algorithms, when varying the normalized channel transmit rate R/G.

The vertical bars represent the standard deviation of the SSIM for the different videos. As expected, SFE and SFC show smaller standard deviation than RF. However, we note that the approximation introduced by the class-based approach is paid in terms of a larger variability of the SSIM of the active videos compared to the exact per-video approach of SFE. Overall, SSIM-based resource allocation schemes tightly reflect the expected behavior of the system for any given SSIM threshold and can further benefit in terms of computational savings when partitioning videos into classes.

Successively, we test the VAC algorithm with the different RMs. To this end, we simulate a Poisson video request process with $\lambda = 0.66$ requests/s. Each video request refers to a video randomly and uniformly picked among the 38 videos of our test set, so that the average offered load is $\lambda T = 10$ videos, where T is the average duration of a video, with an aggregate rate request (at full video quality) of about $G = 150$ Mbit/s. At every new video access request, the VAC invokes the RM algorithm to get the optimal rate allocation in case the new video flow gets accepted. Then, the VAC estimates the SSIM for each active video when applying such allocation policy to the system and checks whether the quality of any video drops below the threshold that we set to $F^* = 0.95$, i.e., the minimum SSIM value to reach a MOS value of 4 (good) (see Table I). In this case, the new video access request is rejected and that video flow is *dropped*, i.e., not activated in the network. When an active video session is over, the related channel resources are released and immediately reallocated by the RM to the active video sequences. Note that, in case of SFC, the class-based SSIM approximation is only considered in the VAC and RM algorithm, while simulations are performed considering the real SSIM characteristics of each video.

Fig. 4(a), Fig. 4(b), and Fig. 4(c) report video dropping probability P_{drop}, average number of active videos, and average SSIM of the active videos, respectively, for the three RM algorithms when varying the rate R of the transmission channel. We observe that when R/G is close to one, the channel rate is large enough to support basically all video requests with full rate, so that the three algorithms perform in a similar manner. When progressively decreasing the channel rate, the RF algorithm scales uniformly down the RSF ρ of all active videos, which determines a *non-uniform* decrease of the SSIM of the different videos. In this way, the SSIM of more dynamic videos get close to the SSIM threshold when others are still enjoying much higher SSIM. Considering again the curves in Fig. 2, we observe that the most dynamic videos reach the SSIM threshold $F^* = 0.95$ for values of ρ from -1.5 to -1, which correspond to a normalized channel rate R/G in the range from 0.02 to 0.1. In these conditions, a single high dynamic active video will prevent the entrance of new videos, since any further reduction of the rate allocated to the ongoing videos will decrease the SSIM of some of them below the threshold. The average SSIM remains approximately constant and equal to the mean SSIM of the videos in the test set for $\rho \in [-1.5, -1]$. For even smaller channel rates, the most dynamic videos are likely not accepted into the system, and the rate of the others will be proportionally reduced, determining a progressive decrease of the average SSIM and increase of the dropping rate.

On the other hand, the SFE and SFC algorithms can admit new videos even at low channel rates (lower dropping probability compared to RF), at the cost of a slight decrease of the quality delivered to the users, though within the constraint of minimum acceptable SSIM. Actually, this constraint cannot be strictly enforced by SFC, whose admission decision is based on the class-based approximation of the SSIM characteristic of a video and, indeed, the average SSIM of active videos is slightly lower for SFC than for SFE. Furthermore, from Fig. 5, we see that the percentage of admitted videos that actually experience an SSIM lower than the threshold is always less than 10%. This non-ideal behavior of SFC, however, can be dealt with by slightly increasing F^*, or making a more conservative choice for the reference SSIM $F^{\mathcal{C}}$ of each class.

VI. Conclusions

In this work, we proposed resource allocation and video admission control algorithms based on the QoE degradation of

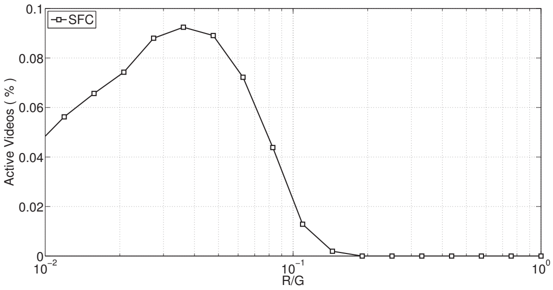


Fig. 5. Percentage of active videos with SSIM lower than threshold because of SFC's class-based VAC policy.

TABLE II
VIDEO TEST SET

Name	Full quality rate [kbit/s]	Duration [s]	Class #
3inrow	11856	12	1
5row1	11135	12	1
Akiyo	5387	10	2
Boblec	11504	12	2
Bowing	10325	10	1
Bridge_close	18246	66	4
Bridge_far	18304	70	4
Vtc1nw	11210	12	1
Bus	16954	5	4
CaesarsPalace	17001	12	3
Cheerleaders	21757	12	4
City	14139	10	3
Coastguard	16570	10	4
Container	12229	10	3
Crew	16179	10	4
FlamingoHilton	25622	12	4
Flower	16335	8	3
Football	15806	3	4
Football_ext	18092	12	4
Foreman	14642	10	3
Hall_Monitor	16291	10	4
Harbour	17929	10	3
Highway	17529	66	4
Husky	24065	8	4
Ice	9517	8	2
Sign_Irene	14091	18	3
Washdc	12948	12	2
Mobile	19172	10	3
Mother_Daughter	11348	10	2
News	7824	10	2
Pamphlet	10917	10	1
Paris	12450	35	2
Redflower	14168	12	2
Silent	11586	10	3
Soccer	14063	10	4
Stefan	17589	3	3
Tempete	17850	8	3
Waterfall	14950	8	3

video clips for slower source rates. We found that, measuring the QoE of a video in terms of SSIM, and expressing it as a function of the logarithm of the source rate, normalized to the full quality rate, it is possible to accurately approximate the quality vs. rate distortion curve of the video as a 4-degree polynomial function. The polynomial coefficients can be tagged to the video and used for QoE-aware resource allocation. To ease the job of RM and VAC algorithms, we further proposed to group the videos with similar SSIM features into classes and to tag each video with its QoE class rather than its SSIM coefficients. We compared the performance of different resource management schemes and we observed that knowing the SSIM vs. rate curve of the videos is beneficial for a real delivery system with strict QoE requirements. Furthermore, the approximation introduced by the class-based tagging of the videos is negligible compared to the performance of the system where the exact knowledge of the SSIM vs. rate curve of each video is given. This is a promising result towards the design of scalable solutions where a network operator has to deal with a very large number of concurrent video flows congesting the links, as forecasted by the latest video growth trend analysis. As a final remark, we observe that the extraction of the SSIM of a video for different encoding rates is computationally expensive. However, from Fig. 2, we observe that certain videos exhibit a relatively large SSIM (> 0.94) even when the transmission rate is decreased to 1% of their full quality rate (e.g., bowing, pamphlet), while others suffer a quick drop of SSIM already when the rate is reduced to 10% of the full rate (e.g., husky, cheerleaders). Videos with similar SSIM behaviors appear broadly homogeneous in terms of scenes dynamics, with most dynamic videos exhibiting a quicker drop of SSIM as the RSF decreases. Thus, it is conceivable to train an automata video classifier that, based on some easy-to-get video features (e.g., frame samples, frame size) can automatically assign the videos to the most suitable QoE class. Our study is a first step towards the design of automated video classification schemes that opportunistically manage the network resources when critical network conditions occur.

VII. ACKNOWLEDGEMENT

This work was supported by the project "A Novel Approach to Wireless Networking based on Cognitive Science and Distributed Intelligence", funded by Fondazione CaRiPaRo under the framework *Progetto di Eccellenza 2012*.

REFERENCES

[1] CISCO, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017," in *White Paper*, February 2013.

[2] N. Amram, B. Fu, G. Kunzmann, T. Melia, D. Munaretto, S. Randriamasy, B. Sayadi, J. Widmer, and M. Zorzi, "QoE-based transport optimization for video delivery over next generation cellular networks," in *IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 2011, pp. 19–24.

[3] Z.Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, pp. 600 – 612, Apr. 2004.

[4] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, pp. 312 – 322, Sep. 2004.

[5] K. Seshadrinathan and A. C. Bovik, "Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos," *IEEE Transactions on Image Processing*, vol. 19, pp. 335 – 350, Feb. 2010.

[6] G. V. Wallendael, S. V. Leuven, J. D. Cock, P. Lambert, R. V. D. Walle, N. Staelens, and P. Demeester, "Evaluation of full-reference objective video quality metrics on high efficiency video coding," in *IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*. IEEE, 2013, pp. 1294–1299.

[7] T. A. Le, H. Nguyen, and H. Zhang, "Scalable video transmission on overlay networks," in *IEEE Second International Conferences on Advances in Multimedia (MMEDIA)*. IEEE, 2010, pp. 180–184.

[8] S. Pasqualini, F. Fioretti, A. Andreoli, and P. Pierleoni, "Comparison of h.264/avc, h.264 with aif, and avs based on different video quality metrics," in *IEEE International Conference on Telecommunications*. IEEE, 2009, pp. 190–195.

[9] "Advanced Video Coding for Generic Audiovisual Services," *ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC.*

[10] L.-Q. Xu and Y. Li, "Video classification using spatial-temporal features and PCA," in *IEEE ICME*, Baltimore, MD, July 2003.

[11] I. Spanou, A. Lazaris, and P. Koutsakis, "Scene change detection-based discrete autoregressive modeling for MPEG-4 video traffic," in *IEEE ICC 2013*, Budapest, Hungary, June 2013.

[12] B. Feitor, P. Assuncao, J. Soares, L. Cruz, and R. Marinheiro, "Objective quality prediction model for lost frames in 3D video over TS," in *IEEE ICC 2013*, Budapest, Hungary, June 2013.

[13] P. Seeling, M. Reisslein, and B. Kulapala, "Network performance evaluation using frame size and quality traces of single-layer and two-layer video: a tutorial," *IEEE Communications Surveys and Tutorials*, vol. 6, pp. 58 – 78, Oct-Dec 2004.

[14] K. Seshadrinathan, R. Soundararajan, A. Bovik, and L. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, pp. 1427 – 1441, June 2010.

[15] T. Zinner, O. Hohlfeld, O. Abboud, and T. Hossfeld, "Impact of frame rate and resolution on objective QoE metrics," in *Workshop on Quality of Multimedia Experience (QoMEX)*, Trondheim, Norway, June 2010.

[16] "Test media repository." [Online]. Available: http://media.xiph.org/video/derf/

[17] "Joint scalable video model - reference software." [Online]. Available: http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm