Exploitation of Producer Intent in Relation to Bandwidth and QoE for Online Video Streaming Services

Michael Riegler[1], Lilian Calvet[1], Amandine Calvet, Pål Halvorsen[1], Carsten Griwodz[1]
[1]Media Performance Group, Simula Research Laboratory & University of Oslo, Norway
{michael, paal, lcalvet, griff}@simula.no, amandine.calvet@gmail.com

ABSTRACT

This paper is the product of recent advances in research on users' intent during multimedia content retrieval. Our goal is to save bandwidth while streaming video clips from a browsable on-demand service, while maintaining or even improving the users' quality of experience (QoE). Understanding user intent allows us to predict whether streaming a particular video in a low quality constitutes a reduced QoE for a user. However, many VoD streaming services today are used by users for a wide variety of reasons, meaning that user intent cannot be inferred from their use of the service alone. However, our investigation demonstrates that user intent does in most cases coincide with producer intent. We can also demonstrate that the latter can be inferred from the content itself as well as associated metadata. By transitivity, we can choose a default video quality that satisfies the users QoE in the majority of cases.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems:**]: [Video]

General Terms

Experimentation; measurement; performance

Keywords

QoE; intent; video streaming

1. INTRODUCTION

Video on-demand (VoD) services like Youtube, Vimeo, Netflix, etc. generate most Internet traffic today. It has been predicted that their share will rise to 90% within the next three years[1]. These on-demand videos are used for a wide range of purposes, ranging from entertainment to education but also communication resembling video mail. Currently, VoD streams are delivered at a default quality chosen by the VoD service provider, independent of their purpose. This implies that a user whose intent it is to enjoy exclusively the music of a music video receives the same video quality as a user who wants to enjoy the sights in a nature documentary. There is a discrepancy since delivering a reduced video quality to users with the first intent would not reduce that user's quality of experience (QoE), for the user with the second intent it would reduce QoE. To make this statement, we do not use the term QoE in the spirit of objective video quality metrics, but rather in terms of the International Telecommunication Union's (ITU) formal definition, which defines QoE as "the overall acceptability of an application or service, as perceived subjectively by the end-user" while "the overall acceptability may be influenced by user expectations and context" [1].

Figure 1: Examples of intent categories. (a) 'Affection': get entertained (e.g. by watching movie); (b)'Experience': learn something (e.g. a recipe); (c)'Information': get informed (e.g. by watching news); (d)'Object': listen to music. Src: Youtube.

We propose a means by which a VoD service can stream videos with a quality[2] that depends on a user's *context* and the *user expectations* to maximize their QoE. For the study conducted in this paper, we restrict the term *context* to typical knowledge of a VoD service provider, such as the user's age, sex and location. We postulate that these simple criteria are sufficient to identify homogeneous user groups whose intent with respect to use of a particular video are likely to be similar. The classification of users by such criteria is

[1]`http://goo.gl/afWfOH`

[2]In this paper, the video quality is expressed in terms of video resolution.

NOSSDAV'15, Mar 18-20, 2015, Portland, OR, USA

http://dx.doi.org/10.1145/2736084.2736095 .

beyond the scope of this paper but they are apparently already exploited by VoD services such as Youtube. Beyond this, *context* includes the situation in which users consume a video stream. Watching news in low quality on a PC monitor in a coffee break may be a satisfactory experience, whereas only a high quality stream satisfies them when watching on a big TV screen at home. The latter challenge has already been explored by analysing user interactions [25].

However, such methods present some limitations: they are mainly designed to determine if the user is interested in both visual and audio content, or the audio content only. If the user is interested in the visual content, the quality that leads to satisfactory QoE may depend on the content itself (e.g. medium for news and high for a movie trailer). User activity may not be sufficient to distinguish these cases.

In this paper, we deal with the specific problem of retrieving the expected quality based on the video content itself. In accordance with our assumptions, we want to establish whether we can deduce QoE from content given the following constraints: (i) *users belong to a single characteristics group*; (ii) *they use the service in the same situation (in their spare time)*; (iii) *they use similar devices (computer with monitor).* We hypothesize that within these constraints, we can select the lowest satisfactory QoE because we can infer the users' intent, i.e. *why* they watch the video, from the content itself. The proposed solution relies on the three following assumptions: (i) Characteristics of a video such as recording, cutting, encoding, etc., have the potential to reveal the **producer intent** so that it is possible to identify producer intent categories based on the video content; (ii) The producer intent reflects the user intent: the main intent of the person who created and uploaded the video and the one of the person who streams it are similar; (iii)Playback quality that provides satisfactory QoE to the user is directly related the the user's intent.

These assumptions modify the interpretation that has been provided by Hanjalic et al. [10]. While we follow the intent categories that they established, namely 'affection', 'experience', 'information' and 'object', which are explained in Figure 1, we do not postulate that user intent is directly connected to video characteristics. Instead, we postulate that characteristics are expressions of producer intent, and that this provides a good prediction of user intent wherever content is consumed as expected by the producer.

The main contribution of this paper is thus to demonstrate the last two assumptions mentioned above. Firstly, we validate the convergence between producers' intent and users' intent. Secondly, we show that, beyond their ability to classify video content, intent categories reveal the default quality that can satisfies the quality expectations of the user. A proof of concept of the proposed system has been developed to validate our assumptions in a user study.

Last but not least, we demonstrate experimentally that the method has potential to reduce the bandwidth considerably for the delivery of some intent categories, while preserving the user QoE. Although the intent computation is quite error-prone (as our experiments also show), it can be used pragmatically if users are allowed to increase quality manually. In such a scenario, temporary dissatisfaction for some users is tolerated, but considerable bandwidth savings can be achieved compared to the alternative always-best-quality approach, while overall satisfaction is higher than in a hypothetical always-worst-default approach.

In Section 2, we outline works related to QoE considerations in distributed multimedia environments, user intent and resource optimization. In Section 3, a conceptual description of the proposed system (illustrated in Figure 2) is provided. Finally, a validation of the above-mentioned assumptions through a proof of concept implementation of the proposed system is described in Section 4.

2. RELATED WORK

Standard internet users are generally not really interested in the technology involved in creating their multimedia content. For most of them, the QoE is the most important concern [12, 11] while watching a video. A lot of research has been done in this direction. For example, Fiedler et al. [9] describe in their work how QoE ties together user perception, experience and expectation to applications and network services. Furthermore, they show how QoE is related to quality of service (QoS).

QoE considerations. In the last years, an increase in the number of distributed multimedia environments, devoting particular attention to QoE requirements, has been observed. At the early stage the issue was that, even if they included user involved interaction, the evaluation of these systems was more system centric. Additionally, the proposed approaches were bothersome for the user, due to the fact that users had to provide additional input. Newer concepts tried to change this direction to a more user centric evaluation based on QoE in combination with QoS [24, 20, 20, 21]. This research ranges from providing a general framework to predicting user QoE. Most of the existing research is based on the network layer and the video encoding/decoding process. Krishnan et al. [16] showed that the quality of the video stream can impact the viewers behaviours. In more detail, they showed that rebuffering and startup time of the video can increase the abandonment rate for a given video.

This is an important insight for our work in combination with the fact that people's major concern is video quality (e.g. in terms of video resolution). So, if we can provide users with the content in a quality that satisfies their needs in terms of QoE, it may give the video provider the opportunity to save bandwidth. We try to tackle this problem by connecting the intent of the video producer with the user intent, i.e. why users want to watch the video. We hope that it can both, help providing the user with a better QoE and help allocating bandwidth in a more adapted way.

User Intent. User intent has been well investigated in research. In particular research has been done in this direction in textual Web search. [23]. Researchers tried to determine what underlying goal the users have when they use a web search engine [3, 6]. Intent has acquired more and more importance in multimedia research in the last years and multiple studies have tried to make the text retrieval approach usable for multimedia [17, 13, 15]. For example, Lux et al. [19] attempted to find possible intent categories for image retrieval similar to the approach presented in [8]. However, these intent-based papers exploit intent in the context of images. With regards to videos, this issue was treated by Kofler et al. [14], who presented an intent ground truth labelled data set. This is important because they show that, as they exist in the context of image retrieval, user intent categories can be identified in context of video retrieval. Hanjalic et al. [10] write about the intent of videos in the context of video retrieval. They present a categorization of videos based on the user intent. Further, they provide a method to classify videos based on their intent, and an evaluation of the classification performance.

A newer approach, called intentional framing [22], looks at the framing of images in order to determine the intent of the photographer by analysing the global visual features of the images. The proposed method is strongly related to this approach as we highlight that how a video is produced (e.g. shooted, mounted, etc.) may reflect the producer intent.

Resource Optimization. In the context of bandwidth awareness, several methods have been proposed such as means to optimize the ratio between energy consumption and bandwidth. [2, 5]. In Microsoft Azure smooth streaming [25], user behaviour and interaction are utilised to adjust the bandwidth usage, e.g. reducing the quality of the video when the video is in the background or displayed simultaneously with another window. Other researchers looked at the potential of analysing video content in order to adapt the bandwidth usage and the video quality. For example, if there are very complex scenes or a lot of movement in the next frames the capacity needed will be higher [18, 4].

Our work differs from current work in the way that we look at the producers intent in correlation with the quality of the video and the quality of the user experience. To the best of our knowledge, the current state of the art does not provide a solution combining intent and video quality in this way.

3. CONCEPTUAL SYSTEM DESCRIPTION

In this section, we describe the general idea and architecture of the system. In order to prove the concept of a multimedia system, able to deliver a content whose quality is related to the producer intent, we implemented parts of the system in a prototype. These parts are described in the experimental section in more detail. The overall system is composed of a client and a server side, shown in Figure 2.

The goal of the proposed system is to understand the user expectations based on the context (if available), the analysis of the video content/metadata and the user behaviour (e.g. via her/his interactions) without requesting any additional information from the user. A typical scenario can be summarized as follows: On the client side, the user is searching for a video while the system is gathering information sent by the user (e.g. text query, url, etc.) on the server side. The playback request then triggers the intent classification. The resulting intent is then considered for deriving a default video quality selection from it. Finally, the video is delivered to the user with respect to the computed quality. If the user is not satisfied with the delivered quality, he/she can change it actively, and any changes in quality settings made by the user is used to feed a semi-supervised machine learning algorithm in order to optimize the expected preferences associated to each intent categories.

Client Side. The first characteristic of the client side is that we apply no or little changes to the standard video player functions provided by video platforms. This way, the client side provides the users with a standard interface similar to those commonly used in video services like Youtube and Vimeo. In this interface, the user is in particular able to change the quality of the video (in the same way as it is provided by Youtube). These quality changes, when they occur, are sent to the system and exploited as expected quality "feedbacks", of which the user is unaware. This information can then be used to adjust the quality setting in relation to the intent category. For example, it is possible that, for a certain intent category, the system does not determine the optimal quality setting but, in this case, this determination could be systematically adjusted based on the overall changes performed by the users. The second information that is collected without awareness of the user is the behaviour of the user regarding the focus of the windows. An example is the video presented in a window in the background and not actively shown, which is strongly connected to the approach from Azure Smooth streaming [25]. This information is particularly useful in detecting certain types of intent categories, e.g., listen to music or a podcast. Information that is not unconsciously provided but a natural input of the user in a video search engine is the search query. Finally, other user-related information can be collected from the user side such as available bandwidth, used display device, etc.

Server Side. On the server side, we implemented so far the intent classification and the intent-to-quality mapping. All other parts are described on a conceptual level. The main system consists of four parts. The first part is the producer intent classification which is responsible for placing each of the videos into one of the intent categories. This is done based on the method presented in [10]. This approach is described in terms of user intent, we adopt it here in order to detect the producer intent. To classify the intent, different sources of information have to be analysed, i.e. the visual and audio content, the metadata and the user input and feedback. The results of this classification will then be combined by late fusion. The second part is the video search engine. This is important because the search query itself can be a valuable source of information about the intent. The third part is the quality association part. This part uses the intent information from the intent classifier to determine the quality of the video delivered to the users. It manages the used codecs but also the final resolution. It creates an intent-quality model that tries to determine the quality of the videos based on the constraint that the bandwidth allocation must be optimized. Furthermore, it also learns from the users feedback (based on whether or not the quality is changed). The last part is the bandwidth thresholding part. This part is responsible for the optimization of the bandwidth usage and is based on the intent and available bandwidth information. It is important to point out that our system will not try to give the user the best quality based on the bandwidth available. It is more a new way to look at the distribution and usage of bandwidth by trying to satisfy the user needs based on the intent without wasting bandwidth.

4. EXPERIMENT

The idea of this initial experiment is to show the convergence between producer intent and user intent, and that different intent categories are correlated to different video quality expectations and therefore bandwidth allocation.

The experiment is split into two parts. The first part is the automatic clustering based on several features of the videos. Since our system is not complete, these initial experiments will show whether building such a system is sensible. The clustering and feature extraction are based on well known methods and frameworks. Moreover, we want to point out that, in future work, we will develop and implement more

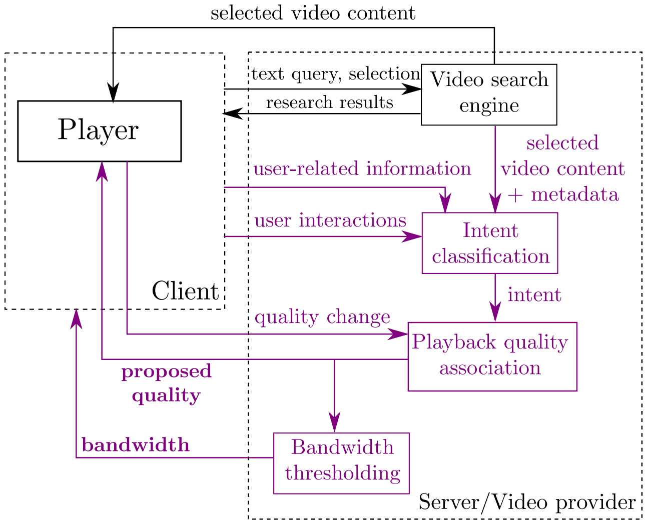

Figure 2: Overview of the proposed system composed of a client side (left) and a server side (right).

sophisticated methods. The second part of the experiment is the user test where we show that the producer intent is correlated to the user intent and that intent is somehow related to the video quality that satisfies the user QoE.

The intent classification part is based on [10]. The classes for the classification are 'Information', 'Experience', 'Affection' and 'Object',i.e. in our context, listen to music. The automatic clustering is performed based on audio and visual features and metadata, which consist of title, description and tags. For the audio information, we used ASR (Automatic Speech Recognition) and for the visual features we used Shot Patterns algorithm. For clustering, we used the Weka machine learning framework [3] and the K-means clustering algorithm. We first calculate the possible cluster for each feature and then combine them in a late fusion step.

For the second part we developed an HTML video player which allowed us to control the quality setting of the video and in the same time collect feedback from the users. We then used a set of 10 trusted users (who we knew would perform the task accurately). Because of the low amount of participants we decided for a placebo-controlled study to make it more to robust. Therefore, we split these 10 users into two equal groups. One group, referred to as by *real group* in the rest of the paper, got videos with quality settings based on the producers intent. The other 5 users, referred to as by *placebo group*, got the same videos with the default level of quality (that we defined as medium with 360p). We chose this way to implement our experiment, in order to, compare the two groups and asses if our method successfully improved the QoE and bandwidth usage combination compared to the standard settings, thus making the experiment more robust.

We downloaded a set of 400 random videos from Youtube that we clustered into the 4 different intent categories. We modified the description of the intents in a way that they are easier to understand for the user. In our case, we chose "listen to music" for the 'Object' intent category. One will maybe assume that music is related to entertainment; this is partially true but music can not be reduced to just entertainment as many people use music for other purposes such as *get relaxed* or *support them at work*. After the classification of the videos we randomly chose 5 videos per intent class. This leaded to a dataset of 20 videos in total for the user test. They range from *cinema trailers* to videos about *how to learn Japanese*. Most of them have a clear intent category. Some can be in more than one category in which case we asked the users for the most fitting one. The video duration varied from some minutes to almost one hour. For the quality representation, we used the Youtube standard settings which are small (240p), medium (360p), large (480p) and hd720 (720p). We did not use higher resolution than 720p because not all of the videos supported it.

[3] http://www.cs.waikato.ac.nz/ml/weka/

Table 1: This table shows the users opinion about the producer intent of the videos in the experiment.

Users Intent / Classified intent	Affection	Experience	Information	Object: listen to music
Affection	46	1	0	3
Experience	2	44	2	2
Information	2	8	38	2
Object: listen to music	11	1	0	38

Table 2: This table depicts the users satisfaction and used bandwidth in MB. Each column presents one intent category (affection, experience, information, object).

Real group				Placebo group				Used Bandwidth in MB			
Preset quality	yes	higher	lower	Preset quality	yes	higher	lower	small	medium	large	hd720
hd720	24	0	1	medium	1	**24**	0	49.31	93.4	141.1	318
large	11	1	**13**	medium	**18**	4	3	22.29	64.98	93.37	265.6
medium	**14**	1	**10**	medium	**12**	6	**7**	22.1	57.2	80.6	201.8
small	**21**	3	1	medium	0	7	**18**	13.26	26.93	38.16	81.4

We then randomly assigned the videos to either the placebo or the real group and each user had to watch all 20 videos and indicate which intent they would choose for each video as well as whether or not they were satisfied with the video quality they were provided with. In introduction to the experiment, a clear and user friendly description of the four intent categories was outlined. In order to insure the correct execution of the assignment, the clarity of the formulation was assessed by performing preliminary tests with five different users. Since we wanted the users to consider the video quality in detail, we formulated the question in a way that arrogates this behaviour. The question for the quality was: *Are you satisfied with the visual quality of the video?*. The possible answers were (i) I would like to watch the video in a higher quality, (ii) I would watch the video in lower quality and (iii) Neither 1 nor 2.

4.1 Results

The collected information support our assumption that producer intent is related to user intent. In consequence, this result suggests that it may be possible to exploit the relation between the producer intent for a video and the quality expected by the user. An overview of the results can be found in Table 1 and 2. The first table contains the opinion of the users from both groups about the producer intent of a video. It can be seen that the users' opinion agrees in majority with the producers intent for the video. The second table shows the user opinions about the quality for each test group and summarizes bandwidth usage in MB per intent class and quality levels for all videos.

Affection. For the affection intent class and in both groups (real and placebo), the participants agreed clearly on the producer intent question. In the group that got the quality settings based on our system the users were satisfied with the quality. They only voted with *yes we are satisfied* or *higher quality*, which we count as satisfied because we set the maximum available quality for the video. In the placebo group, only one user was satisfied with the quality. All other users wanted to watch the video in higher quality, which shows that, the medium quality setting does not satisfy the users quality of experience needs for this intent. In this case, the system uses more bandwidth (compare the last four columns of 2) but the users satisfaction is higher compared to the placebo group.

Experience. For the experience class, we got completely different results as expected. We set the quality for this videos too high (because we assumed that when one experiences something they may want to do it in good quality). For both groups, the opinion about the intent of the videos was clear. The majority of the users in the real test group voted for lower quality. In the placebo group, they were always satisfied with the medium quality (which is one step lower than in the real test group). This gave us two interesting insights. First, the intent of experience is not related with the large quality setting requirements. Secondly, taking the users feedback into account will help improving our system in the future.

Another interesting point was that, one of the videos was an outlier in both groups (affection instead of experience). In the real group, the users were satisfied with the large quality or they expressed the wish for a higher quality. In the placebo group, a higher quality than medium would have been preferred for this particular video. The video was about someone who was playing a computer game and recorded it. This type of videos, called *lets play*, are becoming more and more popular in recent years [4] and are made by the producers for entertainment and not learning purposes. There also exist video platforms which specialize in this type of video [5]. It could definitely also be a video that teaches how to play the game, but such video would have different features regarding content and user-related information.

Information. For the information intent category, we had in both groups a high satisfaction rate. This is justified by the fact that, the medium quality setting was chosen for this intent, which also corresponds to the default setting of the placebo group. Moreover, we had a high precision for the producer vs. user intent classification. An outlier, which was a video about learning Japanese, has been misclassified by our system. This video should be in the intent category of experience/learn something. Another important observation was that, it seems that, user would be satisfied with even a lower resolution than medium for the information intent category. The bandwidth saving potential of this intent category could be even higher.

Object: Listen to Music. The experiment showed us that for this intent category, the lowest playback quality provides satisfying QoE. This can consequently be a very efficient way to save bandwidth without reducing the QoE for the users. An outlier was observed. It was a scene from the *Lord of the Rings* movies, where a Hobbit is singing a song to the lord of a city. Almost all participants voted for this video *affection* as their intent, and they also wanted to see it in a higher resolution, even if most of the part of the clip is a song sung by the Hobbit. We consider this as an indicator that at first, producers intent is very hard to detect. And second, that we definitely need information about the context to be more accurate in the classification part of the system. Finally, the last column in Table 2 reveals, for this intent category, a high potential for saving bandwidth while preserving a playback quality that satisfies the users QoE.

5. DISCUSSION

The experimental results emphasize the convergence between producer intent and user intent. Furthermore, they also agree with the hypothesis that intent information can be exploited in order to adapt the bandwidth distribution. The user votes on the quality satisfaction gave indeed an indication that intent categories are related to quality expectations. It also showed that these categories can help to satisfy the users QoE, either by simply improving it, e.g. with higher video quality, or by preserving it while decreasing the default playback quality. Furthermore, we showed that exploiting these intent information can be a promising idea for interesting bandwidth allocation and saving as it can be seen in Table 2. This can be done based on the fact that videos can be assigned to different intent categories. An allocation based on these intents could help to share bandwidth in a way that the QoE is maximized over all users, or at least, group of users. This would prevent to waste bandwidth by always trying to provide the highest possible quality. This could lead to another important side effect namely saving energy.

The problem of saving energy in the context of online videos has been recently addressed in [7]. The authors looked at the energy consumption caused by different video codecs and video resolutions. A potential problem regarding the applicability of the proposed method is that they have to increase end user awareness and somehow interact with her/him. This can be a challenge as users are generally unwilling when it comes to providing additional information which are not directly associated with their initial goal (streaming a video). Since our system can work independently to the user willingness to cooperate, it could be interesting to further exploit our approach, now for its energy saving potential.

Furthermore, in terms of implementation, it could be very interesting to use DASH [6]. In this scenario, the angle would be changed from just how-much-bandwidth-do-we-have based methods to something more user centred. For example, lets assume the system knows that it has 60 users which want to get entertained, but also 200 users who just want to listen to music. A higher bandwidth (for the better quality) would then only be needed for 60 users who have a need for it, and the rest could be satisfied with the lower bandwidth. This information could be used at the point when the system allocates the bandwidth for the users. It would be an especially interesting alternative when we take the global aspect and the billion of possible users into account and thus its great bandwidth and energy saving potential. This paper is of course just a small step in this direction but the most important insight is the relation between the quality expected

[4] `http://goo.gl/YrvnWf`

[5] `http://www.twitch.tv/`

[6] `http://dashif.org/mpeg-dash/`

by the user and the producer intent, which reflects the user intent in most cases, as a valuable source of information.

A possible limitation of the proposed method is the fact that, it only makes sense in services where the user can search for videos freely like Youtube or Vimeo. Using it in services like Netflix or HBO which have a clear intent before the users start using the service, i.e. in that case *get entertained*, does not seem useful. However, it can not be seen as completely useless because the insights of such a system may be used by these very specialized portals to improve the QoE of their users. For example, it may be interesting to systematically provide low level quality videos on a news video service because the majority of the users will be satisfied with the lower quality level.

Finally, we want to point out that our approach is not only based on the user behaviour or the content. As observed in the experiment section, the user tests showed that users accept lower quality for videos with the intent of information or experience. For these videos, the *being in the background*, *just partially visible* or *just looking at the content* approach would not work well or at all, because it misses the real understanding of the user need. In that case, it makes sense to look at the producer intent. The two approaches are of course complementary and the idea is to adapt the system based on the users feedback but mainly in the sense of learning the intent and the lowest video quality acceptable without impiding a good QoE. Therefore, in a way, we secretly entice the user to using a lower quality without letting them be aware of it. Of course, there will be users that will be unsatisfied and increase the quality but if the main part of the users accept it, bandwidth will still be saved.

6. CONCLUSION

We presented a novel system able to detect a social signal, namely the producers intent and showed that it is related to the users intent for watching a video. We discussed it in context of potential bandwidth and energy saving. The detection of the intent is based on the content, metadata and user related information. Based on the partially implemented system classification, we provided different quality levels to the user. We performed a user study that revealed, that users agree about the producers intent and that they were more satisfied by our system preset qualities than the standard quality setting. This is a strong indicator that such a system can be a new way to look at means to provide content to the users. The next steps include collecting a large scale dataset and conduct experiments over a longer period of time. In future experiments we will also collect information about bandwidth and energy usage levels. This will give us more accurate insight of the possible saving potential.

7. ACKNOWLEDGEMENTS

This work is partly funded by the FRINATEK project ”EONS” (#231687) and the iAD Centre for Research-based Innovation (#174867) by the Norwegian Research Council. We also want to thank Vamsidhar R. Gaddam for discussion.

8. REFERENCES

[1] Vocabulary for performance and quality of service. In *ITU-T Rec.incl. Amendment 2*, 2008.

[2] C. Bae and W. Stark. Energy and bandwidth efficiency in wireless networks. In *Proc. of CCS*, volume 2, pages 1297–1302. IEEE, 2006.

[3] R. Baeza-Yates, L. Calderón-Benavides, and C. González-Caro. The intention behind web queries. In *String processing and information retrieval*, pages 98–109. Springer, 2006.

[4] A. Begen, T. Akgul, and M. Baugher. Watching video over the web: Part 2: Applications, standardization, and open issues. *IC, IEEE*, 15(3):59–63, 2011.

[5] K. Bhardwaj and R. K. Jena. Energy and bandwidth aware mapping of ips onto regular noc architectures using multi-objective genetic algorithms. In *Proc. of SOC'09*, pages 027–031. IEEE, 2009.

[6] A. Broder. A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM, 2002.

[7] O. Ejembi and S. N. Bhatti. Help save the planet: Please do adjust your picture. In *Proc. of the ACM MM*, pages 427–436. ACM, 2014.

[8] R. Fidel. The image retrieval task: implications for the design and evaluation of image databases. *NRHM*, 3(1):181–199, 1997.

[9] M. Fiedler, T. Hossfeld, and P. Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *IEEE NW*, 24(2):36–41, 2010.

[10] A. Hanjalic, C. Kofler, and M. Larson. Intent and its discontents: the user at the wheel of the online video search engine. In *Proc. of the ACM MM*, pages 1239–1248. ACM, 2012.

[11] G. Harman. The intrinsic quality of experience. *Phil. persp.*, pages 31–52, 1990.

[12] R. Jain. Quality of experience. *IEEE MM*, pages 96–95, 2004.

[13] B. J. Jansen, A. Spink, and J. O. Pedersen. The effect of specialized multimedia collections on web searching. *Web Eng.*, 3(3-4):182–199, 2004.

[14] C. Kofler, M. Larson, and A. Hanjalic. First version of an intent ground-truth labeled data set.

[15] C. Kofler and M. Lux. Dynamic presentation adaptation based on user intent classification. In *Proc. of the ACM MM*, pages 1117–1118. ACM, 2009.

[16] S. S. Krishnan and R. K. Sitaraman. Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs. *IEEE/ACM TN*, 21(6):2001–2014, Dec. 2013.

[17] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM TOMCCAP*, 2(1):1–19, 2006.

[18] Z. Li, A. C. Begen, J. Gahm, Y. Shan, B. Osler, and D. Oran. Streaming video over http with consistent quality. In *Proc. of ACM MMSys*, pages 248–258. ACM, 2014.

[19] M. Lux, C. Kofler, and O. Marques. A classification scheme for user intentions in image search. In *CHI'10*, pages 3913–3918. ACM, 2010.

[20] V. Menkovski, A. Oredope, A. Liotta, and A. C. Sánchez. Predicting quality of experience in multimedia streaming. In *Proc. of the AMCM*, pages 52–59. ACM, 2009.

[21] S. Moller, K.-P. Engelbrecht, C. Kuhnel, I. Wechsung, and B. Weiss. A taxonomy of quality of service and quality of experience of multimodal human-machine interaction. In *Proc. of the QoMEx*, pages 7–12. IEEE, 2009.

[22] M. Riegler, M. Larson, M. Lux, and C. Kofler. How 'how' reflects what's what: Content-based exploitation of how users frame social images. In *Proc. of the ACM MM*, MM '14, pages 397–406, New York, NY, USA, 2014. ACM.

[23] D. E. Rose and D. Levinson. Understanding user goals in web search. In *Proc. of WWW*, pages 13–19. ACM, 2004.

[24] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. Yang. Quality of experience in distributed interactive multimedia environments: Toward a theoretical framework. In *Proc. of the ACM MM*, MM '09, pages 481–490, New York, NY, USA, 2009. ACM.

[25] A. Zambelli. Iis smooth streaming technical overview. *MS Corp.*, 3, 2009.