Exploitation of Users Intent for Online Video Streaming Services

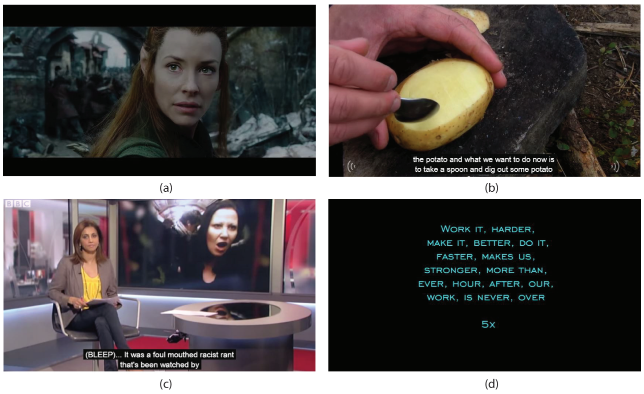
Michael Riegler[1], Lilian Calvet[1], Vamsidhar Reddy Gaddam [1], Pål Halvorsen[1], Carsten Griwodz[1]
[1]Media Performance Group, Simula Research Laboratory AS, Norway
{michael, vamsidhar, paal, lcalvet, griff}@simula.no, mlux@itec.aau.at

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 of 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.

1. INTRODUCTION

Video on-demand (VoD) services like Youtube, Netflix, Vimeo, 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 discrepencay 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" [?]. We propose a means by which a VoD service can stream videos with a quality that depends on a user's *context* and the *user expectations* to maximize their QoE.



(a) (b) (c) (d)

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.

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 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 [30].

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

[1]http://goo.gl/bbYlto

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the author/owner(s).
XXX, XXXX, XXXX, XXXX
ACM XXX-X-XXXX-XXXX-XXXXXXX.
XXXXXX

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:

- 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;
- 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;
- 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. [13]. 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 (presented below) has been developed to validate our assumptions in a user study. Last but not least, we demonstrate experimentally that the method has the 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.

Paper organisation. 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 quality of experience is the most important concern [15, 14] while watching a video. At lot of work has been done in this direction. For example, Fiedler et al. [11] describe in their work how QoE ties together user perception, experience and expectation to applications and network services. Furthermore, they show how quality of experience is related to quality of service.

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 try to change this direction to a more user centric evaluation based on QoE in combination with quality of service (QoS) [29, 25, 25, 26]. 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. [20] 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 work has been done in this direction in textual Web search. [28]. Researchers tried to determine what underlying goal the users have when they use a web search engine based on the information why they use it [2, 6]. "why they use it" Change this part as it does not tie in with the rest of the sentence. This has also been expanded to multimedia retrieval [19, 24]. 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 [21, 16]. For example, Lux et al. [24] attempted to find possible intent categories for image retrieval similar to the approach presented in [10]. However, these intent-based papers exploit intent in the context of images. With regards to videos, this issue was treated by Kofler et al. [17] 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. [13] write about the intent of videos in the context of video retrieval. The authors present a categorization of videos based on the user intent. Further, they provide a method to classify videos based on their intent, and they provide an evaluation of the classification performance.

A newer approach, called intentional framing, looks at the framing of images in order to determine the intent of the photographer [27]. ~~The authors detect the intent that the producer of the content want to convey the content to the consumer.~~ todo: rewrite 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 try and optimize the ratio between energy consumption and bandwidth. [1, 4]. In Microsoft Azure smooth streaming [30], 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 videos 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 [23, 3].

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 architechture 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 side and a server side, illustrated 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, user interactions, etc.) on the server side. The playback request then triggers the intent classification, based on the collected information. 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.

3.1 Client Side

The first characteristic of the client side of our system 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). In order to learn from the user, the system uses unconsciously provided feedback. This is done in several ways.

The first, and invisible to the user, way is the collection by the system of information about whether and, or, and/or ? when the user changes the video quality. This information can then be used to adjust the model for the intent in relation to the quality setting. For example, it is possible that, for a certain type of intent, the system does not determine the optimal quality setting but in this case this determination could be systematically adapted (or readjusted ?) by tracking the quality changes performed by the users. (check major changes to this sentence.) 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, ~~etc.,~~ (in conflict with the fact that you give AN example) which is strongly connected to the approach from Azure Smooth streaming [30]. 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. This information can help determining the intent category of videos which can have more than one possible intent. To explain that more clear, lets take a video of a fashion show as an example. The users could have the intent to *get information* about the newest clothes and trends, but another possible intent could be enjoying the beautiful models (male and female) which can be seen as an *entertainment intention*. The search query for the first intent would then be for example *Colour trends Paris Fashion Show 2014*, and for the second one, *Most beautiful male models at the Paris Fashion Show 2014*. Finally, other user related information can be (or "are") collected from the user side such as available bandwidth, used display device, etc. (check that you are happy with my correction here.)

3.2 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 producer intent categories. This is done based on the method presented in [13]. ~~We adopt this approach to be able to detect the producers intent.~~ (already said above.) 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. All this analysis is very computationally intensive (especially if we take the number of users and videos into account). Therefore, we need to parallelize it ~~in the way,~~ (can it be replace by "meaning" that ..) that we will ~~have distributed~~ distribute classifiers for each input data. The results of these classifiers ~~are~~ 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. What it will also ~~include~~ bring? is the intent as an additional source of information to, e.g., provide intent based search results or recommendations. The third part is the quality estimation part. This part uses the intent in-

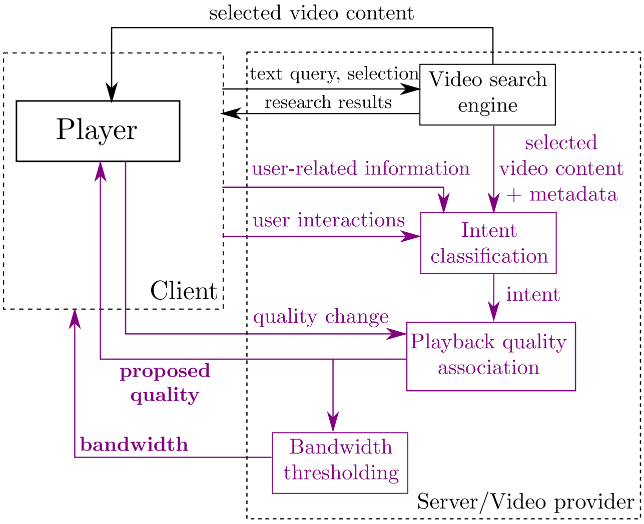


Figure 2: Overview of the proposed system composed of a client side (left) and a server side (right).

formation of the intent classifier to determine the quality of the video presented to the users. It manages the used codecs for decoding and encoding the videos 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. EXPERIMENTS

The idea of these initial experiment is to show that the producers intent somehow reflects the users intent for watching a video and can be used determine the quality. We want to show that the users agree with the producers intent and that different intent categories are correlated to different video qualities and therefore bandwidth allocation, i.e., bandwidth distribution between different users with different intents, how videos are decoded, resolution.

The experiment is split into two parts. Part one is the automatic clustering based on several features of the videos. Since our system is not complete, and these are initial experiments that should show if it makes sense to build such a system, the clustering and feature extraction is based on well known methods and frameworks. We want to point out that in the future work we will develop and implement more sophisticated methods and include them in our own framework as a whole system. The second part of the experiment is the user test where we show that, our hypothesis, i.e., that the user intent is correlated to the producers intent and that the quality of the video is somehow related to them and therefore to the QoE.

The first part of the experiment was the intent classification step based on [13]. The classes for the classification are information, experience, affect and object: listen to music.

For the automatic clustering, we extracted the same audio and visual features and we also used the metadata. The metadata consists of title, description and tags. For the audio information, we used ASR (automatic speech recognition) and for the visual features we applied shot patterns. For the clustering step, we used the Weka machine learning framework [2] and the K-means clustering algorithm. We used late fusion, i.e., we first calculate the possible cluster for each feature and then we combined them. The reason therefore is that we will be able to parallelize it very efficient.

For the second part, i.e., the user experiment, we developed an HTML video player that allows us to control the quality of the video and in the same time to collect feedback from the users. We used a set of 10 trusted users (who we know that they will do the task very accurate.) We split these 10 users into two groups. One group, called real group in for the rest of the paper, got videos with quality settings based on the producers intent. The other 5 users, called placebo group, got the same videos with the default level of quality (that we defined as medium with 360p). The reason is that we can compare these two groups and see if our method really works compared to the standard settings which makes the experiment more robust.

We downloaded a set of 400 random videos from Youtube. Which we clustered into the 4 different intent categories. We modified the description of the producer intents in a way that they are easier to understand for the user. Therefore, we combined the experience classes into one class and we defined the object class by an example. In our case, we decided for listen to music. One will maybe assume that, music is related to entertainment. This is partially true but music can not be reduced to just entertainment. Many people use music to get relaxed or support them at work.

After the classification of the videos we randomly choose 5 videos per intent class. This leaded to a data set 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. We asked the users for the most fitting one. The video duration vary from some minutes to almost one hour.

For the quality representation, we used the Youtube standard setting 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 did support these higher settings.

We randomly assigned the videos to either the placebo or the real group. Each user had to see all 20. Before the users started with the experiment, we gave them a clear and user friendly description about the four intent categories. To find the most clear formulation we performed preliminary tests with five different users to see if the questions are understandable. The task was to tell us which intent they would choose for each video, and we also asked them about the quality of the video and if they are satisfied by it or not.

The question for the intent was: *Please choose a possible intent category for this video based on the description.*

Since we want them to think about the 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

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

Table 1: This table shows the users opinion about the producer intent of the videos used 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

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 user intent is related to the producers intent. Further we showed that makes sense to exploit the relation between the producers intent for a video and the quality. An overview about the results can be found in Table 1 and 2. The first table contains the merged user opinions about the producer intent of a video. It can be seen that the users most of the time agree with the producers intent for the video. The second table show the user opinions about the quality per test group and the summarized 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 (4-5 for affection). In the group that got the quality settings based on our system the users were more 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 us that, the medium quality settings do not satisfy the users quality of experience needs for this intent. In this case the system uses more bandwidth (compare 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 to high (because we assumed that if you experience something you maybe 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 settings. Secondly, taking the users feedback into account will help to improve 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 they stated that, they were satisfied with the larger quality or they wanted higher quality. In the placebo group, they wanted higher quality than medium 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 more and more popular in the last years [3] and made by the producers for entertainment and not learning purposes. There also exist video platforms which are specialized on this type of video [4]. It could definitely also be a video that teaches how to play the game, but such a video would have different features regarding content and by the users provided information. In context of the bandwidth it can be seen that, our system could save more bandwidth if we include the users feedback for learning the optimal quality. It can also be seen that, in this category the possible bandwidth saving potential is promising.

Information. For the information intent category, we had in both groups a high satisfaction rate. This is because of the fact, that we choose medium quality setting for this intent which is the default setting. Further, we had a high precision for the producer intent class classified by our system. 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 insight 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. The results in Table 2 show us that the users most probably would also accept a lower quality than medium.

Object: Listen to Music. The experiment showed us that, for this intent category, the lowest quality setting is the best one. This can be a very efficient way to save bandwidth without reducing the QoE for the users. Interesting was one outlier. 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 user information to be more accurate in the classification part of the system. The last column in Table 2 provides the insight that, also this category has a high potential to save bandwidth and in the same time keep the QoE high.

5. DISCUSSION

The experiment showed us that, the producers intent is strongly related to the intent of the users. Furthermore, it is a good idea to exploit this intent information to distribute and also save bandwidth. The user votes about the quality satisfaction gave us indeed an indication that, intent categories are related to quality. We also saw, that these categories can help to improve the QoE for the users, either by improving it or not reducing it even if the quality is set to a lower level. Further, we showed that it can be a promising idea to exploit these intent information for a substantially bandwidth allocation and saving as it can be seen in 2. This can be done based on the fact that videos

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

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

can be assigned to different intent categories. An allocation based on these intents could help to share bandwidth in a way that the quality of experience is maximized for each user group and in the same moment the bandwidth is not wasted by just trying to provide always the highest quality. This could also lead to another important side effect namely saving energy.

An interesting and new paper about saving energy in context with online videos is [9]. They looked at the energy consumption caused by different video codecs and video resolutions. What we see as a possible problem is that, they have to motivate the users and somehow interact with them. Users most of the time do not really like to provide additional information. which do not really inter associate with their initial goal (streaming a video). Since our system can work without bothering the users to give us additional information it could be interesting for us to additionally exploit the energy saving potential.

For the implementation of our approach it can be a very interesting idea to use DASH [5]. 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 200 users that just want to listen to music. It would then only need around 60 users which have a need for higher bandwidth (for the better quality), and the rest could be satisfied with the lower bandwidth. This could be used as an information for the moment when the system allocates the bandwidth for the users. Especially, when we take the global aspect and the billion of possible users into account and the possible bandwidth and energy saving potential it could be an interesting alternative. This paper is of course just a small step in this direction but the most important thing is that we showed that the quality is related to the producers intent which reflects the users intent in most cases and that this gives us a source of information that is worth to look at it.

A possible weak point for the proposed method is the fact that, it does only make sense in services where the user can search for videos freely like Youtube or Vimeo. Using it in services like Netflix or HBO which somehow have a clear intent before the users start using the service, i.e., in that case *get entertained*, seems not to be useful. But, on the other side, it can not be seen as completely useless because the insights of such a system can be used for these very specialized portals to improve the quality of experience for users. For example, it could be a good idea to provide, from the scratch, 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 just looking at user behaviour or the content. As we observe in the experiment section, the user tests clearly showed us that users also 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 users need. In that case, it makes sense to look at the producers intent. The two approaches are of course complementary and the idea is to adapt the system based on the users feedback but more in the sense of learning the intent for different videos and what is the lowest quality they except for that without having a bad quality of experience. So, in a way, we secretly make them using lower quality without letting them really know. Of course, there will be users that will not be happy and increase the quality but if the main part of the users accept it we can still save bandwidth.

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

6. CONCLUSION

We presented a novel system that is 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 with possible bandwidth and energy saving potential. The detection of the intent is based on the content, metadata and unconscious user information. Based on the partially implemented systems' classification, we provide different quality levels to the user. We performed a user study that revealed, that users agree about the producers intent and that they are more satisfied by our systems preset qualities than the standard quality settings. This is a strong indicator that such a system can be a new additional way to look at the way how we 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 the information about the bandwidth and energy usage level. This will give us more accurate insight in the possible saving potentials.

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.

8. REFERENCES

[1] Vocabulary for performance and quality of service. In *ITU-T Rec. P.10/G.100 (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 *System-on-Chip, 2009. SOC 2009. International Symposium on*, 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. *New Review of Hypermedia and Multimedia*, 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. *Network, IEEE*, 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. *Philosophical perspectives*, pages 31–52, 1990.

[12] R. Jain. Quality of experience. *IEEE MultiMedia*, 11(1):96–95, Jan. 2004.

[13] B. J. Jansen, A. Spink, and J. O. Pedersen. The effect of specialized multimedia collections on web searching. *J. 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 Trans. Netw.*, 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. *Microsoft Corporation*, 3, 2009.