Systematic Assessment of the Video Recording Position for User-generated Event Videos

Stefan Wilk
Distributed Multimedia Systems
TU Darmstadt, Germany
stefan.wilk@cs.tu-darmstadt.de

Wolfgang Effelsberg
Distributed Multimedia Systems
TU Darmstadt, Germany
effelsberg@cs.tu-darmstadt.de

ABSTRACT

The increasing capabilities of mobile handhelds to record high definition videos enable us to share moments with friends and the public. In large-scale events such as concerts, the manifold of potential views recorded by the audience allows to watch an event from different perspectives. In this work, we elaborate on the effects of different recording positions on quality and show that especially with retail smart phone camera sensors the quality differs as the position changes. This is a first systematic approach to describe why user-generated video sequences differ in terms of quality depending on changing recording positions. To achieve a sound understanding, we conduct a large-scale systematic subjective study using crowdsourcing, inspecting the effect of distance and orientation of a recorder in relation to the event taking place. Our results indicate an effect of the recording position on the perceived quality of a video.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Evaluation/ Methodology

Keywords

User-generated Video; Crowdsourcing; perspective; distance; orientation; smart phone; Video quality; QoE

1. INTRODUCTION

Sport events or music festivals motivate hundreds in the audience to save precious moments as digital video. Video sharing sites, such as YouTube, profit from the fact that those clips are uploaded and shared with the public. Many of the recorded video clips show the same scene but recorded from different positions. The question thus arises in what ways the recording position affects the perceived quality for a potential watcher. The knowledge about its relation to the quality can help to tune recommender systems of video sharing sites. In addition, video mashup systems, which allow to automatically direct a video mix from different recording cameras, and the recording users could benefit from this knowledge.

This work elaborates on different positions and angles for recording an event. To conduct an in-depth empirical analysis on the effect of the position on the perceived quality, we had to record several sequences in the genres sports, entertainment and scenery from different positions at the same time. We have conducted experiments on a crowdsourcing mediating platform and validated results in a lab setting. The sequences were evaluated by 451 assessors.

Video quality can be determined using objective video quality metrics. Especially full-reference methods such as VQM by Pinson [3] result in values with high correlation to the subjectively perceived quality. But at the same time they require high processing times. No-reference methods exist that enable real-time computation, e.g. Yang's approach [4], but with reduced precision. However, none of these metrics so far can estimate the effect of different positions in the recording phase of a video creation process as they were designed to identify compression and network artifacts. Crowdsourcing is a new technique in multimedia research that allows us to perform a large-scale, subjective user study at low cost. Crowdsourcing is based on the design of small and well-defined tasks that can be completed in minutes. These tasks are provisioned to an anonymous group of people using a mediating platform. In contrast to regular subjective lab tests the evaluation can be easily distributed to a less biased and at the same time larger group of users. This advantage comes with the drawback of anonymous assessors conducting the tests in a remote location with reduced control of the test settings. Properties and best practices for conducting video quality assessments using the crowd are well described in the work of Hossfeld [1].

The contribution of this work is the systematic evaluation of recording positions on the perceived quality in different live performances. Our work illustrates a significant negative impact of a increasing distance to the recorded live performance, whereas differing recording angles show no significant differences.

2. APPROACH

This section discusses the setup of our crowdsourced video quality assessment to evaluate the effects of different positions while recording a scene in video. In a first step, different video sequences had to be recorded in order to create

MM'14, November 3–7, 2014, Orlando, Florida, USA.
Copyright 2014 ACM 978-1-4503-3063-3/14/11 ...$15.00.
http://dx.doi.org/10.1145/2647868.2655022.

a dataset that contains enough material to be evaluated on large-scale. The qualitative assessment is based on crowdsourcing using a web-based evaluation system. Our crowdsourcing experiment asks different unknown users, in the following called assessors, to watch multiple video sequences of the same event. The task of the assessor is to judge on the perceived video quality of each sequence using the Single Stimulus Continuous Quality Scale (SSCQS) recommended by the ITU [2]. Tasks are clustered in so called campaigns. Each campaign represents a distinct set of video sequences from one of the genres: 'sports', 'entertainment' or 'scenery'. Each scene is recorded from different distances and angles which resulted in 8 evaluated events in 79 sequences. The sequence order has been chosen randomly for each assessor which in combination with the large number of assessors reduces the effect of biasing the subjective ratings by a fixed viewing order. In total 451 assessors assessed the sequences which resulted in 3160 ratings. The video clips are evaluated in regard of the distance to the event and the angle. Zero degree represents a frontal view to e.g. the stage as depicted in Figure 1. Only valid angles and distances were chosen depending on the genre of the recorded event. This means e.g. for recording a concert far beyond one hundred meters will not result in a sensible recording quality using standard smart phones. These clips have been discarded. Another example is a 'scenery' sequence such as the Eiffel tower in Paris for which distances of several hundred meters will result in good video quality. Those sequences are included in the dataset. Thus, we cluster the distance into: close-up, medium, medium-long and long shot. In addition, we list for each scenario the distances that are associated with the different category levels. The angles of the orientations are measured in 10 degree steps from the origin with 0 to 90 degrees. The assumption is that beyond the 90 degrees angle no sensible recording can be created. In those cases, harmful occlusions occur due to the setup of a classical, proscenium show stage. For central staging, e.g. in a stadium an rectangular stage was assumed and each side is evaluated on its own.

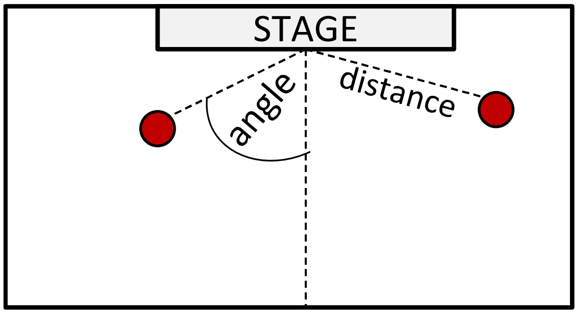


Figure 1: Illustration on measured attributes of a position: distance and angle

The videos that were basis for this evaluation were recorded during sports events (a soccer game and motor sports) different entertainment events such as concerts or shows, and points of interest in different cities ('scenery'). Entertainment and sports videos have been recorded during live performances in Darmstadt and Frankfurt, Germany. Entertainment videos include a circus comedy event (Entertainment 1), an artistic performance including rapid movements due to jumps (Entertainment 2) and a music concert with a crowded audience. Sports events recorded include a soccer game (Sport 2) and a motorbike competition including jumps. All recordings show the same region of interest, but with different level of detail. Scenery sequences have been taken in Paris, France showing the Eiffel tower in different views (Scenery 1) and in Darmstadt, Germany of a historic building (Scenery 2) and representative statues (Scenery 3). Table 1 lists the recorded events and the preferred recording distances.

All video sequences have a duration between 9 and 12 seconds. The audio tracks are removed from the videos. The video sequences are lossy compressed during recording and have a resolution of 4CIF. The camera has been set up to ensure low quantization rates during encoding. Lossy encoded videos are used as those are predominantly used in UGV. For the crowdsourcing as well as the lab tests we ensure that the video clips could be played continuously without stalling. Examples of the videos are shown in Figure 2.

The lab experiment was conducted with 15 test subjects based on the recommendations of the ITU [2]. This included a random selection of test subjects, a test for viewing impairments of the subjects, proper lighting conditions, room and technical settings. A training session in the lab setting and a qualification test in the crowdsourcing group have been held but no data from the training session or qualification test is used in our final results. Subjective quality assessments but especially crowdsourced evaluations suffer from the lack of reliability of the assessors. A data cleaning step based on the recommendation of the ITU [2] has been conducted. In addition, the valid ratings received from the assessors may show very biased results as many assessors tend to rate in the lower regions of our 5-point SSCQS scale, whereas others will more likely give high scores. To compensate this issue we normalized the gathered ratings. Normalization of the ratings calculates the value r'_{ij} of a assessor i and video j from the original rating r_{ij} corrected by an adjustment of the mean rating of an assessor $\overline{r_i}$ by the mean of the campaign $\overline{r}$, where each campaign represents one recorded event with several sequences.

$$r'_{ij} = r_{ij} - (\overline{r_i} - \overline{r}) \tag{1}$$

All ratings are aggregated to the Mean Opinion Score (MOS) by:

$$MOS_j = r_j = \frac{\sum_{i=1}^{N} r'_{ij}}{N} \tag{2}$$

3. RESULTS

The results describe the distance to a recorded event as well as the angle under which it was recorded.

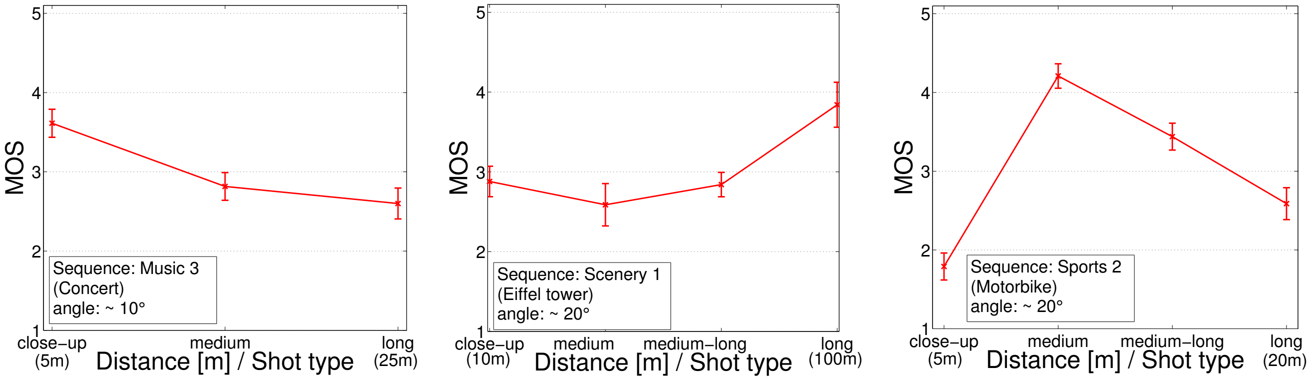
3.1 Impact of the Distance

The impact of distance to the observed event may be one of the major factors for decreased visibility and thus reduced quality. Figure 3 shows the distribution of the perceived quality for varying distances using the same orientation. It shows that the distance has an observable effect for different recordings.

For the 'entertainment' genre, in this case a concert, our results indicate that the close-ups are seen as a preferable shot type for short music recordings. In this case close-up means, that a limited part of the stage is shown, and the fo-

Sports 2 Entertainment 3 Entertainment 1

Figure 2: Impressions of our video dataset used to evaluate the perceived quality based on the recording position.



(1) Entertainment 3 - Concert (2) Scenery 1 - Eiffel tower (3) Sport 2 - Motorbike

Figure 3: Perceived quality of recordings from different distances with the same orientation for (left) Entertainment 1, (center) Scenery 1 and (right) Sports 2.

cus lies on the main performer, omitting the assisting band. The video was recorded only some meters away from the stage, not using any internal software-assisted zoom functionality of the camera. From the close-up to more distant shot types only a minor quality degradation can be observed (see Figure 3 - left). Distances from 5 up to 30 meters are preferred in stage performances. For the entertainment genre it can be observed that close-ups and medium shots or in general: close distances to the stage - result in best qualities. Distances beyond 30 meters lead to a significant drop in perceived quality due to the evaluated frame size of the videos and the technical limitations of smart phone camera sensors.

In addition, the assessors were asked to annotate the Region of Interest (RoI). Although the annotated RoI in most cases showed a significantly larger region of the scene, the sub-region building the close-up was preferred in the concert clips. A more significant drop of the perceived quality was observed for the sequences in the genres 'Scenery' as well as 'Sports'. In case of the 'Scenery' sequences, e.g. sequence 'Scenery 1' (see Figure 3 - center) the perceived quality increases with the distance at least until the point when the whole RoI can be seen in the sequence. It seems that scenery recordings include a wider border region around the RoI in comparison to the entertainment sequences, which indicates that the viewers are more interested in the surroundings of the point of interest. For the sports recordings the medium shots show the main actor, e.g., the soccer player leading the ball or the motorbike rider performing stunts. This distance is preferred in comparison to close-ups or far distant overview shots. This indicates that a preferred shot type for our user-generated sport clips is recorded in a medium shot distance, allowing a combination of overview as well as close connection to the central actions in the game. The figure as well as the evaluations of the other sequences (see Table 1) allow us to conclude that a sweet spot for a distance can be determined depending on the genre of the video. Within the genres, the specific event recorded shows only small deviations for the preferred shot type.

Table 1: Description of the different events evaluated in the experiments.

Event	Content	Preferred shot
Entertainment 1	Circus show	medium (15 m)
Entertainment 2	Artistic show	medium (15 m)
Entertainment 3	Concert	close (5 m)
Scenery 1	Eiffel tower	long (100 m)
Scenery 2	Historic building	long (60 m)
Scenery 3	Statue	medium (17 m)
Sports 1	Soccer	medium (15 m)
Sports 2	Motorbike	medium (11 m)

3.2 Impact of Recording Angle

The orientation, i.e., from which angle a scene is recorded, builds the second factor we evaluated in this work. The perceived quality represents similar quality levels in all genres for angles between 0 and 70 degrees. Significant differences in the perceived quality may result from some little differences

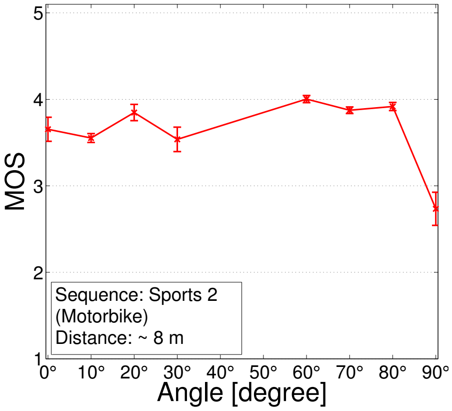


Figure 4: Perceived quality in different orientations for the observed scene 'Sports 1'

Table 2: Correlation of the MOS between crowdsourcing and lab experiments.

Genre	Pearson	Spearman
Music	0.7137	0.7692
Show	0.9854	1
Scenery	0.9375	0.8571

in the recordings that cannot be avoided in non-synthetical data. In extreme cases, especially with 80 to 90 degrees or more, it can be observed that the quality drops significantly (see Figure 4). We conclude that approaches for identification of the orientation of a recording in relation to the observed event should limit their effort on the identification of extreme angles as they may limit the perceived quality. In contrast to that, our results indicate that videos recorded at angles common in events with clear separation of stage and audience areas can result in good to excellent quality. Similar results were found in other genres which indicate that the orientation, at least in non-professional recordings is not as important as other factors. In future work, we will elaborate on the effect of orientations of different recordings in long-running sequences of several minutes. Our expectation is that the importance of diversity in long-running videos will boost the effect of different orientations.

4. CORRELATION

Although the reliability of our workers is ensured in our crowdsourced evaluation applying the methods mentioned in Section 2, we still had to ensure that the ratings retrieved in a remote, uncontrolled setting correlate with controlled lab settings. A lab test based on ITU-R BT.500-13 [2] with 15 test subjects was conducted for the sequences in the three genres. For the purpose of demonstrating a correlation the Pearson's R coefficient as well as the Spearman Rank was chosen. The Pearson's coefficient calculates how much the points scatter around a linear trend, whereas the Spearman Rank shows whether a monotonic function can describe the relationship of the crowd and the lab datasets. In Table 2 the Pearson's coefficient as well as the Spearman Rank for the different video groups can be retrieved.

Especially for the genres 'Sports' and 'Scenery' a high correlation between the perceived quality in the remote crowdsourced experiments can be observed. The 'Entertainment' sequences show slightly differing results especially for the reduced quality for increasing distances. In comparison to the crowdsourced tests increasing distances are perceived degrading the quality significantly more. Due to the low number of lab participants, we believe the crowdsourced results to be more reliable.

5. CONCLUSION

Mobile handhelds capable of recording video allow all of us to share experiences during events such as concerts or sport games with the world. On video sharing sites, multiple videos of the same event are uploaded. Visitors of such sites do not know the quality of the recordings. Many different attributes affect the quality of such a video clip. One of them is the recording position. This work describes the impact of video recording positions on the perceived quality in UGV. It is the first systematic crowdsourced, subjective analysis of the impact of different positions on the perceived quality in user-generated video. The results indicate a strong influence of the distance on the perceived quality. In contrast to this, the orientation seems to have a limited impact on sequences recorded by non-professionals. Additionally, we show that the preferred view differs depending on the recorded scene and event. By using the results of this work, recommender systems for video sharing sites can be designed in a more efficient way - ranking the results of a search depending on the recording position. Our next step includes the investigation whether these findings in combination with position data gathered from smart phone sensors can be used to perform optimal view selection for video summary and video mashup algorithms.

Acknowledgments

This work has been funded by the German Research Foundation (DFG) in the Collaborative Research Center (SFB) 1053 'MAKI: Multi-Mechanism-Adaptation for the Future Internet'.

6. REFERENCES

[1] T. Hossfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia. Best practices for QoE Crowdtesting: QoE assessment with Crowdsourcing. *IEEE Transactions on Multimedia*, 16:541–558, 2014.

[2] ITU-R. BT.500: Methodology for the subjective assessment of the quality of television pictures (ITU-R BT.500-13). Technical report, International Telcommunications Union, 2012.

[3] M. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 50(3):312–322, 2004.

[4] F. Yang, S. Wan, Y. Chang, and H. Wu. A novel objective no-reference metric for digital video quality assessment. *IEEE Signal Processing Letters*, 12(10):685–688, 2005.