iCaCoT - Interactive Camera-based Coaching and Training

Lucia D'Acunto
TNO
Anna van Buerenplein 1,
2595DA, The Hague,
The Netherlands
lucia.dacunto@tno.nl

Judith Redi
TUDelft
Mekelweg 4,
2628 CD, Delft,
The Netherlands
J.A.Redi@tudelft.nl

Omar Niamut
TNO
Anna van Buerenplein 1,
2595DA, The Hague,
The Netherlands
omar.niamut@tno.nl

ABSTRACT

This paper reports on the evaluation of the concept of *interactive camera-based coaching and training* (iCaCoT), which focuses on using interactive video navigation for coaching and training purposes. The iCaCoT concept leverages *tiled streaming* technology, which allows users to navigate freely through high-resolution video feeds while minimising the bandwidth required, by only streaming the part of the video the user is interested in.

iCaCoT gives a trainer the possibility to zoom in on her trainee while she is training and to focus on specific areas, both spatially and temporally. This concept becomes especially useful for training activities where the exact line followed by the trainee is not known beforehand (e.g. skiing, footballing), and thus where capturing the events using a static wide-angle camera located relatively far from action may be more convenient than a moveable close-up camera. We implemented the iCaCoT concept as an iPad application and demonstrated it with ski athletes in the popular ski location of Schladming, Austria. Our experiment shows that iCaCoT is a viable concept for ski training activities and that it gives interesting insights for future research directions.

Author Keywords

interactive video navigation, adaptive streaming, tiled streaming, coaching and training, quality of experience, experiment, experimental research.

ACM Classification Keywords

H.5.1. Information Interfaces and Presentation: Multimedia Information Systems - evaluation/methodology

INTRODUCTION

With the advent of high resolution and panoramic cameras, which are able to record in HD or higher resolutions, it becomes interesting to segment content spatially. By dividing a video frame up into multiple *tiles*, where each tile contains a particular area of the video, a client can choose to only receive certain areas of a video. Such a tiled streaming solution enables an inherently scalable method for users to interact with and navigate within a video using pan-tilt-zoom (PTZ) commands. In the EU FP7 project FascinatE [1] we have implemented such a *tiled streaming* technology in an iPad application to enable users to navigate freely through high resolution video panoramas, while the application limits bandwidth requirements by only sending that part of the video a user is interested in.

The concept of tiled streaming looks particularly well suited to training and coaching use cases. That is, using a smartphone or tablet, a coach would be able to zoom in on her trainee while she is training, focusing on specific areas, both temporally as well as spatially. We refer to this as *interactive camera-based coaching and training* (iCaCoT). This concept becomes especially useful for training activities where the exact line followed by the trainee is not known beforehand (e.g. in skating, skiing, footballing, baseballing), and thus where capturing the events using a wide-angle camera located relatively far from action may be more convenient than a moveable close-up camera. By pausing the video at key moments, trainer and trainee can focus on and discuss details of the performance. By placing multiple high resolution cameras around strategic positions, it is even possible for a trainer to view a moment from different angles. The tiled streaming application facilitates this using high-accuracy synchronization techniques, ensuring that the separate videos from all cameras are synchronized frame-accurately in the application.

In this paper, we present the results of an evaluation of the iCaCoT concept with ski athletes, performed between February and March 2014 at the popular ski location of Schladming, Austria (host of the 2013 Alpine Skiing World Championship). Conducting an experiment with real users has enabled us to study and evaluate the suitability of tiled streaming as a tool for coaching and training in practice and understand the key enablers for interactive camera-based coaching and training. Specifically, we were interested in answering the following research questions:

1. What are the relevant aspects for a camera-based coaching and training application?

2. Is iCaCoT a suitable tool for training and coaching activities?

As a subquestion of the second question, we were also interested into understanding the overall user experience when interacting with the iCaCoT application.

3rd International Workshop on Interactive Content Consumption at *TVX'15*, June 3, 2015, Brussels, Belgium.

To answer these questions we have collected and analyzed a number of metrics ranging from application features usage to Quality of Experience (QoE) parameters.

RELATED WORK

With recent capturing systems for high-resolution video, new types of video-based training scenarios are possible where trainers and coaches have the possibility to freely choose their viewing direction and zooming level. Different examples of such interactive region-of-interest (ROI) video streaming have already been demonstrated or deployed. Interactive ROI video streaming was explored in-depth by [6, 7]. The authors developed various methods in the context of an interactive ROI streaming system, ClassX, for online lecture viewing, selecting *tiled streaming* as the best compromise between bandwidth, storage, processing and device requirements. Tiled streaming relies on a tiling of video into independently decodable video streams. Client devices retrieve the tiled videos corresponding to a desired ROI. A similar zoomable video system was further explored by [10]. There, the focus was on enabling low-delay interaction with high-resolution and high-quality video, with constraints on the available bandwidth and processing capabilities as encountered in current network technologies and devices. For the iCaCoT application, we leveraged the tiled streaming system and mechanisms as presented in [8, 12].

In today's sport training and performance analysis, nearly all performances are captured on video or through other sensors. Captured footage and sensor data is then viewed by expert coaches/analysts, who then manually annotate and label important performance indicators to gauge performance. Related work in sport performance analysis ranges from reducing annotation time [11], to computer-assisted self-training systems for sports exercise [5], extracting tactic information next to regular semantic event detection [13], leveraging virtual reality for a better understanding of the many biomechanical, physiological, and psychological factors [4], and using on-body acceleration sensors to perform motion and flying force analysis of ski-jumping [3].

In this paper, we provide novel contributions by focusing on a trainer's user experience when interacting with a training application. We present the results of an initial QoE evaluation of an interactive camera-based coaching and training application based on tiled streaming, performed "in the wild". We further investigate important application functionalities and QoS of the underlying operational live video tiling system. The scale and complexity of the field trial makes these contributions very relevant for assessing the business opportunities of the interactive video system and training application.

DESCRIPTION OF THE ICACOT SYSTEM

This section details the overall architecture of the iCaCoT system, including backend, frontend and monitoring framework (Figure 1).

Backend

For the experiment, we designed and developed a pipeline for a live tiling system consisting of the following components:

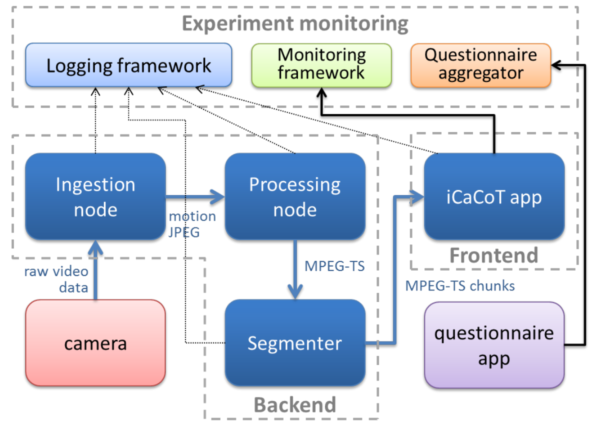

Figure 1. High-level architecture of the iCaCoT system.

- *Ingestion node*, which captures the raw video data and encodes it using Motion JPEG codec;
- *Processing node*, which takes the input encoded in Motion JPEG received from the Ingestion node and produces viewable video files; this step includes the tiling, encoding (in H.264/AVC) and multiplexing (MPEG-TS container) of the content;
- *Segmenter*, which produces the temporal segmentation of the content, i.e. the final streams, using Apple's HLS solution.

The output of the Segmenter is subsequently distributed to the different instances of the iCaCoT app via a webserver.

Frontend

The frontend of the iCaCoT system has been implemented as an iOS app for iPad. In addition to the distinguishing functionalities of tiled streaming (pan and zoom in/out), a number of additional ones have been included in the implementation of the iCaCoT app, to fit the purpose of coaching and training. These functionalities can be broadly classified into two categories:

- **GUI functionalities**
  - *Pan*: to navigate within the high resolution video stream;
  - *Zoom in/out*: to change the level of details by switching between different resolution representations of the video stream;
  - *Draw*: to draw lines as overlays on the video stream;
  - *Bookmark*: to store a certain position in time (a maximum of 6 bookmarks can be stored);
  - *Pause/resume*; to pause and resume the video stream;
  - *Step-frame*: to step through frames when the video playback is paused; a trainer can use this function to show an athlete her exact moves and explain what to improve;

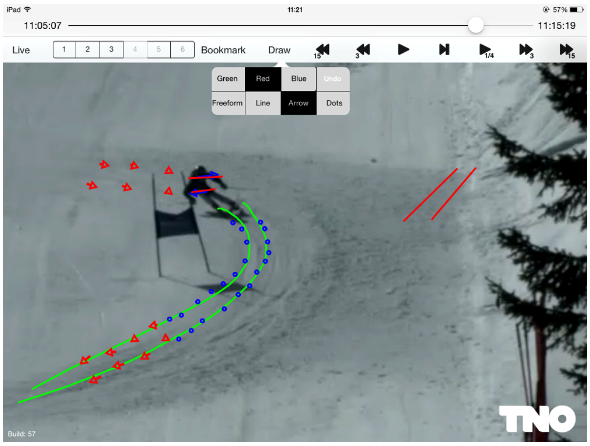

Figure 2. iCaCoT application screenshot. After a trainer has zoomed into a specific region of interest and has paused the video, he has used the drawing functionality to highlight critical training aspects, such as posture and tracks.

- *Seek*: to move playback to another point in time with respect to the current position; (to +3, +15, -3, -15 seconds); this function can be used to look for a specific point in time (e.g. a particular athlete's movement).

Furthermore, for the second experimental run (see the section "Experimental Setup") the following additional GUI functionalities have been added:

- *Enhanced draw functionality*: line-based, arrow-based and dot-based drawing, plus the ability to choose different colours;
- *Slow motion playback functionality*: plays the video at 1/4 of the original speed.

Experiment monitoring

Throughout the experiment, we have been monitoring app usage, user experience, network parameters and system components' behaviour through a monitoring framework. The monitoring framework comprises the following:

- *Monitoring framework*: to monitor network and application usage. We have used EXPERIMonitor, a baseline component of the FP7 project EXPERIMEDIA [9], for:
  - *Network monitor*: data downloaded over time and missed video frames (collected every ms);
  - *Usage monitor*: every user interaction with the iCaCoT app - pause/resume draw, seek, zoom in/out, pan (collected at event occurrence);
- *Logging framework*: to collect real-time info from each component in the architecture in Figure 1; this information is used for debugging purposes.
- *Questionnaire aggregator*: to collect the trainers' subjective evaluations of iCaCoT. The questionnaire aggregator is a part of the QuickTapSurvey tool [2], which also included a questionnaire app deployable on all ipads; the information collected was completely anonymous.

Figure 3. Reiteralm location for the iCaCoT experiments (Schladming, Austria). The three cameras are represented in red and the cabin hosting our equipment in green.

EXPERIMENTAL SETUP

To provide answers to the research questions mentioned in the introduction, we have conducted a number of experiments with ski trainers in the popular ski location of Schladming in Austria. This section outlines system deployment and experiment description for our study.

System deployment

We chose the Reiteralm area as setting for the experiments, because of it being well-suited for (semi-)pro coaching and training purposes. Figure 3 shows an overview of the slope used in our experiment (slope 3), including locations of the three cameras used during the experiments and the cabin hosting our backend and monitoring equipment. Cameras location and orientation were agreed upon with the ski trainers.

System setup involved various challenging tasks, such as installing and connecting cameras over distances of hundreds of metres on the skiing slope, and installing cables via underground bunkers (Figure 4).

Experiment description

Over a period of 2 months, we have performed two experiment rounds. The first took place in week 8 (February 17-21) and the second in week 13 (March 25-29) of 2014. Each experiment round saw the participation of 4 trainers, each testing the app with a group of 7-10 athletes. For each ski athlete, a trainer would use the app for two key activities: (i) watch each athlete live as he/she is coming down the slope, and (ii) discuss with each athlete his/her performance using playback of the recorded video. Before the start of the experiment, trainers were briefed over the functionalities of the

Figure 4. Impressions from system setup.

iCaCoT application, especially on those specifically designed for training purposes (zoom, pan, bookmark, trickplay). Once the experiments started, our experimenters were closely monitoring the execution, reminding the trainers about the available functionalities and advising on their usage. After each trainer concluded his training activity, he was asked to fill in a questionnaire (via the questionnaire app on the iPad) about his experience with iCaCoT.

The first experiment round was used, among others, to gain insights into the needs of the end users (the ski trainers): how they envision using the app and what features they require. Using the information obtained in the first experiment round, we have made improvements for the second round. Improvements included advanced GUI functionalities (as described in the section "Description of the iCaCoT system") and higher resolution cameras (from the GoPro of the first round to a Blackmagic Design 4K camera of the second round).

Collected data

To evaluate our experiments, we have collected both objective data (network and usage monitor measured by the iCaCoT app), and subjective data (through questionnaires). These are discussed in detail below.

Objective data

Throughout the experiments, the iCaCoT app logged a number of usage and network metrics from participants. Every minute, the app would send the data collected in the last minute to the EXPERIMonitor. The metrics being logged included the current bitrate, the total data usage, the region of the video that a trainer was viewing, the app feature being called, and a dropped frame during playback. In our analysis, we only considered *droppedFrame* and *featureCall*.

Subjective data

We have used questionnaires to assess the ski trainers' impressions of the iCaCoT app.

The questions have been divided into the following categories:

Table 1. Occurrences of each app functionality per minute across the two experimental rounds. p-values refer to a U-test of the data.

Feature	Median (# 1)	Median (# 2)	U	p-value
pause	0.66	0.49	350	0.1879
resume	0.54	0.24	353.5	0.1637
seek forward	1.28	3.10	253.5	0.4464
seek backward	0.53	1.28	240	0.3284
pan	6.53	3.88	356	0.1502
zoom	36.97	9.54	458	0.0003866
draw len	11.28%	3.09%	372	0.07035
step-frame	42.54	0	194	0.0004414
add bookmark	0	0	383	0.0151
select bookmark	0	0	310	0.4283

- User satisfaction: measures user quality of experience by asking the trainer direct questions (e.g. whether their experience was good or bad)
- GUI usability: measures whether the trainer can interact fluently with the app. This includes two aspects:
  - Ease of learning, which measures whether the trainer intuitively learn how to use the app
  - Ease of interaction, which measures whether the application features have been implemented in the correct way
- Functional usability: measures whether the application features work as they should (e.g. no major hiccups within application usage)
- Application value: measures whether the trainer perceives that the app is useful for his/her training activities

Based on the feedback from the first round, we could determine that a number of questions were less relevant for the trainers (such as the ones on the enjoyability or friendliness of the app) or for the second round (e.g. the ease of learning and interaction, since the trainers were already used to the app). Therefore these questions were removed in the questionnaire presented at the second round. Furthermore, a few new questions have been added in order to assess the impact of changes/improvements done to iCaCoT between the first and second round. A comprehensive list of questions can be found in Table 2.

We obtained and analysed 7 questionnaires in total (4 filled in during the first round and 3 filled in during the second).

EXPERIMENT EVALUATION

Relevant aspects for training

Our first research question aims at investigating what the relevant aspects of a camera-based coaching and training systems are. To answer this question we have analyzed the subjective evaluations and the app usage. From the *UsefulFeature* open question, the slow motion functionality appeared to be the most popular (40% of the respondents), followed by step-frame and draw (30% of the respondents each). This result was expected, as the slow motion functionality was added after the first experimental round upon trainers' feedback.

Each user could open and close the app several times during the same experiment. We will refer to the app usage

Table 2. **Overview of the items included in the questionnaires adopted in the two experimental rounds. All questions (excluding Yes/No questions) are on a 5-points scale.**

Question/Variable	Round	Scale	Abbreviation	Category
Experience with the app	1 and 2	ACR	Experience	User satisfaction
Expectations with respect to the app	1 and 2	Bipolar	Expectations	User satisfaction
Enjoyment of the app	1	Agreement	Enjoyment	User satisfaction
Excitement of the app	1	Agreement	Excitement	User satisfaction
Endurability of the app	1 and 2	Yes/No	Recommendation	User satisfaction
Ease of learning the app	1 and 2	Bipolar	Learnability	GUI usability (learning)
Ease of understanding the app	1	Bipolar	Understandability	GUI usability (learning)
Friendlyness of the app	1	Bipolar	Friendliness	GUI usability (interaction)
Ease of use of the app	1 and 2	Bipolar	Usability	GUI usability (interaction)
Predictability of the app during usage	1	Bipolar	Predictability	GUI usability (interaction)
Comprehensiveness of the app functionalities	1 and 2	Bipolar	Comprehensiveness	Functional usability
Performance improved with respect to exp 1	2	Bipolar	PerfImprovement	Functional usability
Ability to follow athlete skiing live	1 and 2	Agreement	QualityLive	Functional usability
Ability to find playback of athlete skiing	2	Agreement	Searchability	Functional usability
View over the piste from the app	1	Bipolar	CameraPositioning	Functional usability
Smoothness of video navigation	1 and 2	ACR	QualityNavigation	Functional usability
Video quality of the app	1 and 2	ACR	QualityVideo	Functional usability
Video quality improved with respect to exp 1	2	Bipolar	QualityImprovement	Functional usability
Dissatisfaction with interruptions	1 and 2	Yes/No	Interruptions	Functional usability
Satisfaction with startup delay	1	Bipolar	StartupSatisfaction	Functional usability
Satisfaction with the latency of the live video	1 and 2	Bipolar	LatencySatisfaction	Functional usability
Overall continuity of the video stream	1 and 2	ACR	Continuity	Functional usability
Usefulness of the app	1 and 2	Bipolar	Usefulness	Application value
Innovativeness of the concept	1	Bipolar	Innovativeness	Application value
Impact on teaching ability	1	Bipolar	TeachingImpact	Application value
Impact on students' learning curve	1	Bipolar	LearningImpact	Application value
Impact on teaching time	1	Bipolar	TeachingTime	Application value
Beneficial for trainers	2	Yes/No	TrainerBenefit	Application value
Beneficial for athletes	2	Yes/No	AthleteBenefit	Application value
Most useful app feature	1 and 2	-	UsefulFeature	Open
Change in the app	1	-	AppChange	Open
Remove from the app	2	-	AppRemove	Open
Add in the app	1 and 2	-	AppAddition	Open
How to improve playback search	2	-	SearchImprovement	Open

within consecutive opening and closing as a "session" and analyze app usage parameters per session. Eventually, we recorded parameters for 26 usage sessions during the first experiment and 22 during the second, across all participants. From the recorded data, we have then calculated the number of occurrences of each functionality per minute for each session. We wanted to verify whether these app usage statistics were significantly different across the two experimental rounds, possibly as a consequence of the change we made in the system (enhances GUI functionalities and higher quality camera). For this purpose, we used a non-parametric Mann-Whitney U-test, which checks whether the medians of two (non-normal) distributions are equal. Table 1 reports the median values for each app functionality usage in both session, the test statistic and the significance value (p). As we can observe, pan and zoom are among the features that were used most frequently per minute, with a median of 6.53 and 36.97 in the first round and 3.88 and 9.54 in the second round, respectively. Step-frame also scored high in the first round (median 42.54 times per minute). It was almost never used in the second round, probably a consequence of having introduced slow-motion (which the trainers used for the same purpose of illustrating the details of a certain movement to an athlete). Furthermore, we have also calculated the fraction of time during each session that a trainer spent drawing. From Table 1 we note a trend in that trainers spent less time drawing in the second round, compared to time spent in first round. This might be due to the enhanced drawing functionality provided in the second round, but given the p-values of this U-test this assumption needs further investigation.

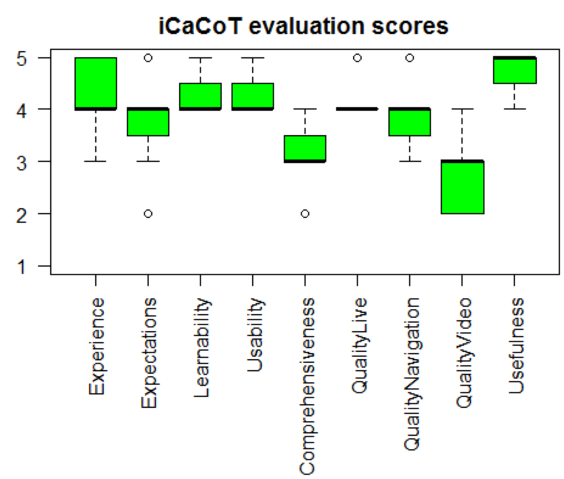

Figure 5. Overall evaluation of iCaCoT from the questionnaires

iCaCot for coaching and training and user experience

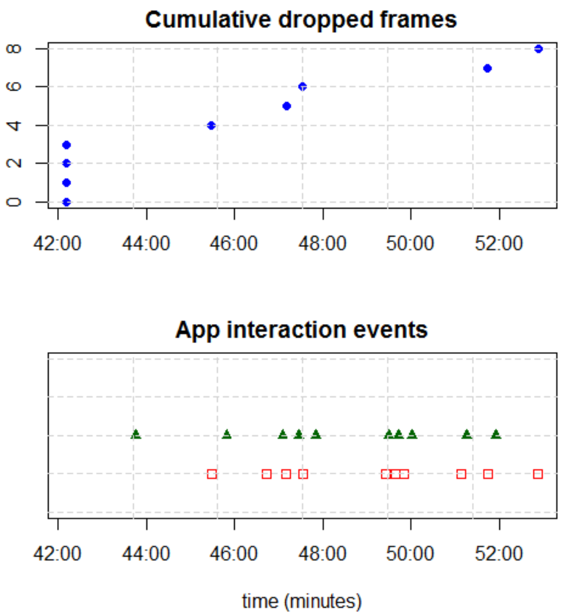

Figure 6. Cumulative dropped frames (top figure) and app interaction events (bottom figure) during a session run. The green triangular points represent *pause* events while the red square points represent *resume* of playback events.

With our second research question, we seek to understand whether the iCaCoT concept is suitable for coaching and training activities. Figure 5 shows the scores given to iCaCoT across both experiments. As we can see, iCaCoT scores high for experience, learnability, usability, quality of navigation and, most importantly, for usefulness. This trend is also reflected in evaluation of impact on teaching/learning ability and benefit for trainers and trainee. Furthermore all participants would recommend iCaCoT to others. These results show that ski trainers found the app very valuable, which is an indication that iCaCoT is a viable concept for coaching and training.

On the other hand, iCaCoT scored a bit lower on the quality of the video and comprehensiveness of the functionalities (Figure 5) and 43% of the respondents indicated that they noticed too many interruptions in the video feed (which is also confirmed by a median of 17 dropped frames per minute). We noticed that frames were dropped during certain trick play events, such as seek or resume playback (See Figure 6. Additionally, trainers have indicated (in the open question about additions to the app) that they would further benefit from a method for tracking athletes, comparing 2 athletes or 2 runs of the same athlete and a timer. Further research on tiled video streaming for the use in coaching and training should focus on these aspects.

CONCLUSIONS

This paper presented the implementation and results of an experiment "in the wild" with an interactive camera-based application for coaching and training. Although conducted on a small scale, the results of our experiment provide indications that this type of applications are in fact very valuable for both trainers and trainees. Additionally, thanks to a combination of network data, app usage data, and subjective evaluation from the participants, we were able to identify a number of relevant aspects that affect the experience and satisfaction of trainers with the iCaCoT concept. For example, we noticed a trend about some network parameters (dropped frames) being related to app usage and we believe that further studies should focus on exploring these relationships in more detail. We also observed that trickplay and draw functionalities are of paramount importance for ski trainers. Nevertheless, improvements can still be made to the functionalities made available from the app, especially for what concerns tracking ski athletes and visualization of different training performance at the same time. Further research in this domain should focus on these challenges.

REFERENCES

1. Fascinate. `http://www.fascinate-project.eu/`. Accessed: 2015-03-16.
2. Quicktapsurvey.com. `https://www.quicktapsurvey.com/admin/import/`. Accessed: 2015-03-16.
3. Bachlin, M., Kusserow, M., Troster, G., and Gubelmann, H. Ski jump analysis of an olympic champion with wearable acceleration sensors. In *Wearable Computers (ISWC), 2010 International Symposium on* (Oct 2010), 1–2.
4. Bideau, B., Kulpa, R., Vignais, N., Brault, S., Multon, F., and Craig, C. Using virtual reality to analyze sports performance. *Computer Graphics and Applications, IEEE 30*, 2 (March 2010), 14–21.
5. Chen, H.-T., He, Y.-Z., Chou, C.-L., Lee, S.-Y., Lin, B.-S., and Yu, J.-Y. Computer-assisted self-training system for sports exercise using kinects. In *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on* (July 2013), 1–4.
6. Mavlankar, A., Agrawal, P., Pang, D., Halawa, S., Cheung, N.-M., and Girod, B. An interactive region-of-interest video streaming system for online lecture viewing. In *Packet Video Workshop (PV), 2010 18th International*, IEEE (2010), 64–71.
7. Mavlankar, A., and Girod, B. Spatial-random-access-enabled video coding for interactive virtual pan/tilt/zoom functionality. *Circuits and Systems for Video Technology, IEEE Transactions on 21*, 5 (2011), 577–588.
8. Niamut, O., Prins, M., van Brandenburg, R., and Havekes, A. Spatial tiling and streaming in an immersive media delivery network. *Adjunct Proceedings of EuroITV* (2011).
9. Phillips, S., B. M. B. M. C. S. E. V. S. Z. W. S. Linking quality of service and experience in distributed multimedia systems using prov semantics. *Service Oriented System Engineering (SOSE), 2015 IEEE 9th International Symposium on. l* (2015).
10. Quax, P., Issaris, P., Vanmontfort, W., and Lamotte, W. Evaluation of distribution of panoramic video sequences in the explorative television project. In *Proceedings of the 22nd international workshop on Network and Operating System Support for Digital Audio and Video*, ACM (2012), 45–50.
11. Sha, L., Lucey, P., Morgan, S., Pease, D., and Sridharan, S. Swimmer localization from a moving camera. In *Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on* (Nov 2013), 1–8.
12. van Brandenburg, R., Niamut, O., Prins, M., and Stokking, H. Spatial segmentation for immersive media delivery. In *Intelligence in Next Generation Networks (ICIN), 15th International Conference on*, IEEE (2011).
13. Zhu, G., Xu, C., Huang, Q., Rui, Y., Jiang, S., Gao, W., and Yao, H. Event tactic analysis based on broadcast sports video. *Multimedia, IEEE Transactions on 11*, 1 (Jan 2009), 49–67.