The Cameraman Operating My Virtual Camera is Artificial: Can the Machine Be as Good as a Human?

VAMSIDHAR REDDY GADDAM, RAGNHILD EG, RAGNAR LANGSETH,
CARSTEN GRIWODZ, and PÅL HALVORSEN, Simula Research Laboratory
and University of Oslo

In this article, we argue that the energy spent in designing autonomous camera control systems is not spent in vain. We present a real-time virtual camera system that can create *smooth* camera motion. Similar systems are frequently benchmarked with the human operator as the best possible reference; however, we avoid a priori assumptions in our evaluations. Our main question is simply whether we can design algorithms to steer a virtual camera that can compete with the user experience for recordings from an expert operator with several years of experience? In this respect, we present two low-complexity servoing methods that are explored in two user studies. The results from the user studies give a promising answer to the question pursued. Furthermore, all components of the system meet the real-time requirements on commodity hardware. The growing capabilities of both hardware and network in mobile devices give us hope that this system can be deployed to mobile users in the near future. Moreover, the design of the presented system takes into account that services to concurrent users must be supported.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Video*

General Terms: Experimentation, Measurement, Performance

Additional Key Words and Phrases: Interactive immersion, panorama video, zoom, panning, real-time, visual servoing, virtual camera, quality of experience, user studies

ACM Reference Format:
Vamsidhar Reddy Gaddam, Ragnhild Eg, Ragnar Langseth, Carsten Griwodz, and Pål Halvorsen. 2015. The cameraman operating my virtual camera is artificial: Can the machine be as good as a human? ACM Trans. Multimedia Comput. Commun. Appl. 11, 4, Article 56 (April 2015), 20 pages.
DOI: http://dx.doi.org/10.1145/2744411

1. INTRODUCTION

Improvements in communications and processing power provide opportunities to explore groundbreaking systems in interactive immersive applications. For example, in scenarios like surveillance and sports, high-resolution wide field-of-view panoramic video has become popular. Stitched panorama videos generated from a camera array that covers an entire field can be used to support virtual views through zoom and pan operations. In turn, individual users can interactively control their own virtual camera. We have created a prototype system that provides an immersive experience using a soccer stadium as a case study. The strength of the system lies in complete automation of several steps that are currently considered to be superior when operated by a human. In this respect, a user can function as his or her own cameraman. However, in

This work has been performed in the context of the iAD center for Research-based Innovation (project number 174867) funded by the Norwegian Research Council.
Author's address: V. R. Gaddam; email: vamsidhg@ifi.uio.no.

ACM 1551-6857/2015/04-ART56 $15.00
DOI: http://dx.doi.org/10.1145/2744411

many situations it may be desirable to allow the system to automatically operate the cameras, for instance when following the ball or a particular player in a soccer scenario.

Most commonly seen broadcast is usually of high profile games. In such games, the financial resources can make several simultaneous high quality capture systems possible. However, most soccer games at different levels of profession do not qualify for such high production qualities. In such cases, there are typically 2-3 simultaneous video streams that are switched. Considering a few games,[1] it can be observed that the camera operator majorly works with pan-tilt-zoom operations for capturing the game from a single viewpoint. Thus, our main research question in this article is whether a machine-generated virtual camera can provide a viewing experience that is at least as good as a human-generated one, under similar conditions?

Many researchers have looked at similar challenges [Ariki et al. 2006; Carr et al. 2013; Chen and De Vleeschouwer 2010] from a broad point of view, but focussing on different details of such systems. Our overall goal is to facilitate personal interaction rather than consider the users as passive observers, which is typically the case in traditional uni-stream broadcasts. This personal interaction could involve the manual control of the virtual camera or the decision to allow automatic tracking of objects. One way to provide an interactive presence in the stadium is to deliver video in which a user can pan, tilt and zoom from a given viewpoint, the position of the cameraman. Ideally, if a user was present at the stadium, these camera movements are the degrees of freedom he or she should have without moving. However, when scaling such a system to several users, the delivery part is constrained to be independent from the capturing part (no physically moving cameras based on user needs). In this article, we briefly show how the panorama is generated and how the virtual view is extracted from the panorama. Then, we present different ways to automatically control the zoom, pan and tilt. Finally, we perform a 2-step user study where the first step focuses on comparing different algorithms for servoing the virtual camera and the second step evaluates the machine generated movements against those generted by human operators. We analyse the user studies aiming to answer the question whether a machine has the potential to be as good as a human operator, and the answer is promising.

The remainder of the article is organized as follows. Section 2 briefly outlines some of the many related works in relevant intersection areas. Then, Section 3 introduces our system, before we look into the details of the automatic camera control algorithms in Section 4. In Section 5, we present experimental results on the technical implementation, and Section 6 encompasses the user studies that evaluate manual and automatic camera controls. Finally, we discuss the results and implications in Section 7 before we conclude the article in Section 8.

2. RELATED WORK

Our system contains many integrated components. In this section, we briefly describe the ones we found most similar and closely relate to our approach.

Ren et al. [2010] provide details about an 8-camera system that is able to keep track of players using hypothesis from the multiple views. However, the scope of their paper is limited to extracting the position information. Several free-viewpoint systems [Carranza et al. 2003; Debevec et al. 1996; Grau et al. 2004; Kanade et al. 1997] exist to provide the viewer with the power to change the view point to the desired one smoothly. However, all those have limited challenges due to the fact that they are made indoors. Outdoor sports provide ample number of challenges to reuse the same techniques in terms of space, illumination changes and uncontrolled conditions. Thus

[1] http://www.youtube.com/watch?v=E8jSOv8Ch5s - [10:00 - 11:00]
http://www.youtube.com/watch?v=FEM0dY8c0co.

the depth based image rendering techniques [Papadakis et al. 2010; Grau et al. 2007] still widely suffer to achieve the production quality. This can be seen in the recent work by Goorts et al. [2014]. However, impressive the functionality is to a researcher, the visual quality is still far from delivery to general audience. Hence we looked at single view-point approach.

First, generating a panorama video has by itself several challenges, and extensive amounts of literature is available for panorama creation. A well-known example is the panoramic capture system that was used to record a world cup match during FIFA World Cup 2006 [Fehn et al. 2006]. Similar works include [Xu and Mulligan 2013; Gaddam et al. 2014a; Carr and Hartley 2009] where the authors emphasize on engineering challenges related to a system for recording panoramic videos. Nevertheless, recent approaches prove that panorama video can be generated in real time [Gaddam et al. 2014b], then the challenge remains to extract virtual views from the panorama.

Using a panoramic texture for an immersive feeling is also a researched topic [Jenkin et al. 1998; Chen 1995], also in different applications areas (e.g., lecture videos [Yokoi and Fujiyoshi 2005; Sun et al. 2005; Ahmed and Eades 2005]). Some works also describe manually controlled virtual cameras generated from panoramic videos with emphasis on human-centered interfaces [Foote et al. 2013], network performance [Mavlankar and Girod 2010] and over-all system aspects [Gaddam et al. 2014b].

Automation of camera control at several levels has also been explored before. For example, Wang et al. [2004] provided a multilevel framework to automatically generate replays from just one camera. Dearden et al. [2007] provide an evolving system that learns from the movement of a trained camera operator. Several works focused solely on control theory of virtual cameras from multiple cameras [Lipski et al. 2009; Christie et al. 2005; Hutchinson et al. 1996]. Though these are interesting approaches, we want to build an entire system that extracts a virtual view from a high-resolution panorama controlled either by the user or a machine operated cameraman in real time. The idea is to put together these different components and bridge the gap so as to make these components function in coherence.

We are definitely not the first to explore such ideas. For example, Ariki et al. [2006] provided a prototype where they used clipping on a portion of an HD recording as a means of creating a virtual camera. They generate the virtual camera motion automatically based on situation recognition from the game. Their user study focuses more on learning the effect of various evaluation criteria like naturality in zooming, panning, shot size, duration, video quality and intelligibility on the audience preferences. This is probably the closest and a simpler version of our work. Furthermore, Carr et al. [2013] presented a hybrid system using both a robotic PTZ camera and a virtual camera generated from panorama. They evaluated their system comparing it to a human operated one as benchmark. Their motivation is to get as close to the human operator as possible. Even though a really thorough work dealing with automatic virtual cameras, they fixed the focal length and the tilt angle subjecting to less exposure to scrutiny by the viewers about the short-comings from changing these variables. Similarly, Chen and De Vleeschouwer [2010] performed an automatic production planning over multiple cameras. They employed an individual stimulus rating based evaluation system which cannot be directly used for comparing different variables. Both of these works focus on basketball as their case study, which has a much more limited field size compared to a soccer field, giving smaller panning requirements.

However, the main focus of these investigations is whether the perceived experience from a automatically controlled virtual view can match the one generated by a human. We try to learn from the existing approaches and design a system for automatic control of a virtual view extracted from a high-resolution, wide field-of-view panorama video

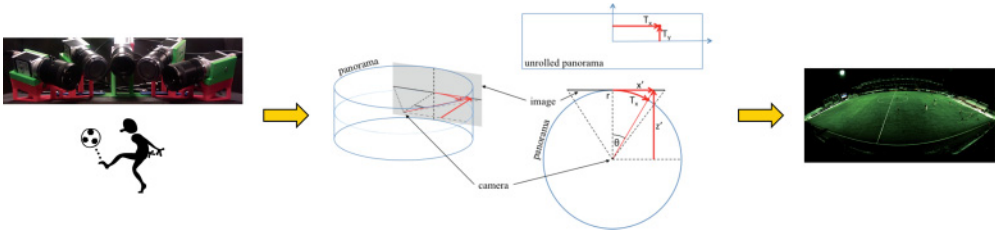

Fig. 1. The camera array captures videos from the individual cameras, each with its own spatial coverage. The idea is to have the cameras in the center of a virtual cylinder where every source image is considered a (cropped) plane, tangential to the cylinder and orthogonal to the camera's viewing axis. Each pixel of the cylinder is then computed as the (interpolated) pixel value of the ray from the camera center through the pixel intersecting the image plane.

and evaluate it. But we do not use the human operator as a benchmark, instead as a competitor in our evaluations.

3. SYSTEM SETUP

Our solution is composed of a sensor system to capture player positions, a mobile-phone-based expert annotation system to tag events live, and a camera-array-based video system to generate a wide field-of-view panorama video in real time [Halvorsen et al. 2013]. The current prototype is deployed and running in an elite club stadium.

Because the main focus of this article is the video system, we first describe the generation of the panorama video before we describe how the system enables individual users to control their own virtual camera.

3.1. Panorama Recording System

To capture the visual sequences and generate a real-time panorama video [Tennøe et al. 2013], we use a distributed recording system. Five Basler cameras capture videos at a 1086 × 2046 pixel resolution, arranged as illustrated in Figure 1. Recording machines thereafter transfer the captured raw video data over a PCIe-based high-speed interconnect network to a panorama processing machine. To give a higher vertical resolution, the cameras are rotated 90 degrees. Moreover, the frequent changes in the outdoor light conditions require auto-exposure to be performed on the center camera at fixed intervals, with the resulting exposure parameters broadcasted to the other camera readers. The model in Figure 1 portrays how the processing machines warps the frames onto a cylinder and then stitches the cylindrical panorama, with the seams calculated dynamically for each frame in order to avoid the ghosting artifact of the moving objects (the players and the ball). The resulting stitched videos have output frames of 4096 × 1680 pixel resolution, as seen on the right of Figure 1. Finally, the processing machines encode the panoramic video in H.264, making it available for HTTP streaming.

3.2. Virtual View Generation

A pinhole camera model is applied to generate a virtual view [Gaddam et al. 2014b] of the field of interest, with pixels fetched from the panoramic texture. The reprojection onto the virtual camera plane preserves a perspective nature of the output view, where the straight lines remain straight, irrespective of any distortions introduced during stitching. The perspective nature is accomplished using the pin-hole-based point projection from a 3D point P to an image point q, which can be written as follows:

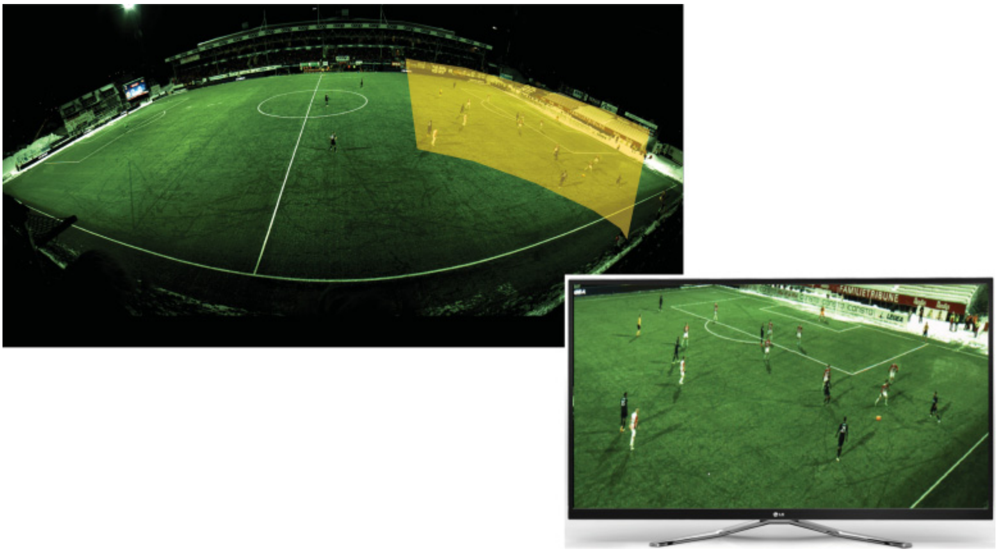

Fig. 2. Virtual view generated from re-projection. The panorama video with the marked region of interest is shown together with the generated virtual camera, emphasizing that the extracted area is not a simple crop from the high-resolution panorama video.

$$\lambda q = [K|0_3] \begin{bmatrix} R & 0 \\ 0_3 & 1 \end{bmatrix} \begin{bmatrix} 0_3^T & -C \\ 0 & 1 \end{bmatrix} P, \tag{1}$$

where R is the general (3 × 3) 3D rotation matrix as a function of θ_x, θ_y and θ_z, the rotation angles around the x, y and z axes, respectively. Moreover, K is the camera intrinsic matrix built with focal length (f). Then, if p is the current pixel, we need to find the ray (s) that passes from the camera center C to the pixel p:

$$s = \lambda R^{-1} K^{-1} p. \tag{2}$$

Then, the intersection point of this ray with the unit cylinder gives us the exact position on the cylindrical texture:

$$T_x = \left(\frac{W_p}{FOV}\right) \left\{ arctan\left(\frac{-s(1)}{s(3)}\right) \right\} + \frac{W_p}{2} \tag{3}$$

$$T_y = \left(\frac{1}{2} - \frac{s(2)}{\sqrt{s(1)^2 + s(3)^2}}\right) H_p. \tag{4}$$

Here, the point (T_x, T_y) represents the coordinates on the unrolled cylindrical texture as described before, and W_p, H_p and FOV correspond to the width, height and field-of-view of the panoramic texture, respectively. When these calculations are performed with subpixel accuracy, the intersection will not necessarily land at one pixel. Consequently, an interpolation may be required from the surrounding pixels, which we manage using bicubic interpolation [Gaddam et al. 2014b]. Depending on the requested output resolution, the entire virtual camera frame is generated in about 10 ms. An example of a generated view is included in Figure 2.

In the current setup, the client controls and generates the virtual view using a client program. This program fetches the decoded video segments, before the final virtual view is controlled either manually or guided by the orchestrating program. The latter is achieved using position data, in this scenario, the position of the ball or the selected players, with either object tracking [Kaiser et al. 2011; Xu et al. 2005; Yu et al. 2003] or available sensor data.

In the current context, we have applied automatic tracking and we describe in this article key approaches for virtual camera movements. By doing this, we seek to minimize computational expenses, thus enabling the client to run on devices with different capabilities, ranging from high-capacity desktop machines to mobile phones.

4. APPROACHES FOR AUTOMATIC CAMERA CONTROL

To operate the virtual camera automatically, we are operating in 3D space with the origin on the axis of the cylinder for all the movements. Here, we let θ_x be the angle along the pan direction, θ_y be the angle in the tilt direction, and f be the focal length, that is, these three variables are used and changed to control the virtual camera. A ray pointing at $(\theta_x, \theta_y) = (0, 0)$ meets the panorama image at the center.

Furthermore, let the feature point on the panorama be $s_p = (\theta_x^p, \theta_y^p)$, and let the current state of the camera be $c^i = (\theta_x^i, \theta_y^i, f^i)$ where previous states are denoted $c^{i-1}, c^{i-2}, \ldots$. Then, the problem of operating the virtual camera can be formulated as

$$c^i = F\left(s_p^{i+l}, s_p^{i+l-1}, s_p^{i+l-2}, \ldots, c^{i-1}, c^{i-2}, \ldots\right), \tag{5}$$

where l is the future data fetched by simply delaying l units of time. The broadcast can be slightly delayed from real-time and this is quite a common phenomenon with delays attributing to delays in channel, direction process etc. However, the processing in our system is automatic and strictly real time, we can introduce an artificial delay to provide some future data to the servoing algorithms. This helps us keep the causality of the system, because in reality, the future data has already been captured. The models that we developed for controlling the virtual camera handle the state variables independently. There are two models for controlling the angles and the focal length is controlled depending on the current position of the center of the virtual camera on the panorama.

4.1. Models for Pan and Tilt

We have used two different models for the pan/tilt operations, that is, a Schmitt trigger and an Adaptive trigger. The pan and tilt angle movements are assumed to be independent. However, the changes in tilt angles are penalized more than the pan angles because panning is usually more natural than tilting a camera in wide field of view situations.

4.1.1. Schmitt Trigger. The concept of a schmitt trigger is to stabilize noisy input data. We modified it so as to provide a smoother movement by adding an acceleration α. For the schmitt trigger to function, we define an imaginary window[characterized by θ^t] inside the virtual view. When the target point is inside the imaginary window, the system is brought quickly, yet smoothly, to rest by using an acceleration α_{stop}. Once the target point goes outside the window, we provide an acceleration α, to the view so that we reach the target. The sign of α depends on the current velocity of the feature point and the virtual camera. The acceleration is added only when the velocity is less than the maximum velocity $\delta\theta_max$. Algorithm 1 presents this approach. The velocity and acceleration of a variable θ are written as $\delta\theta$ and $\delta^2\theta$, respectively.

4.1.2. Adaptive Trigger. The adaptive trigger is designed to adaptively estimate the required velocity of the virtual camera. We smooth the movement of the camera in a

ALGORITHM 1: Schmitt Trigger

```
if $\theta^p$ is outside $\theta^t$ then
    if $\delta\theta^p > \delta\theta^{i-1}$ then
        $\delta^2\theta \leftarrow \alpha$
    else
        $\delta^2\theta \leftarrow -\alpha$
    end if
else
    $\delta^2\theta \leftarrow \alpha_{stop}$
end if
```

two step smoothing process. We use a running weighted mean smoothing at both steps. Another key difference in this model is the use of future data. By delaying the system by 1 second, we have future data for about 1 second. The windows for the regression are smaller than the fetched future data because of the second level smoothing. For a given variable x let $S(x)$ be the smoothed value. Algorithm 2 describes this approach. When computing the target velocities, the gradient is taken over smoothed feature positions because the noise get amplified with a gradient. τ is a threshold for removing small variations in position that are caused by small jerky motions. These jerky motions create a small average velocity over multiple frames. We preferred to keep the camera static rather than subjecting it to a really slow movement.

ALGORITHM 2: Adaptive Trigger

```
$(\theta_x^0, \theta_y^0) \leftarrow (\theta_x^p, \theta_y^p)$
while running do
    $\delta\theta^s = \delta(S(\theta))$
    if $\delta\theta^s > \tau$ then
        $\delta\theta^{st} = \delta\theta^s$
    else
        $\delta\theta^{st} = 0$
    end if
    $\delta\theta = S(\delta\theta^{st})$
end while
```

4.2. Models for Zoom

The zoom is controlled by modifying f accordingly, the virtual view is zoomed by increasing f. In the current system, we developed two models to change f depending on where the virtual view is looking at. With our knowledge from different broadcasts, we wanted to find out the preference of audience for these two commonly used zoom mechanisms.

4.2.1. Smooth Zoom. This is to imitate the nature of the physical zoom that is obtained by smoothly controlling the zoom ring on the recording camera. We modelled a quadratic function in the current camera position coordinates such that f increases when the position approaches the goal posts or the other end of the field from the camera setup.

$$f^i = \lambda_0 + \lambda_1\left(\theta_x^i - \theta_{x0}\right)^2 + \lambda_2\left(\theta_y^i - \theta_{y0}\right)^2, \tag{6}$$

where λ_1 and λ_2 are the parameters that control the effect of pan and tilt angles respectively. θ_{y0} is used to offset the curve so that the function is an increasing one over all the tilt angles. θ_{x0} is set to 0, because the function should be increasing from the

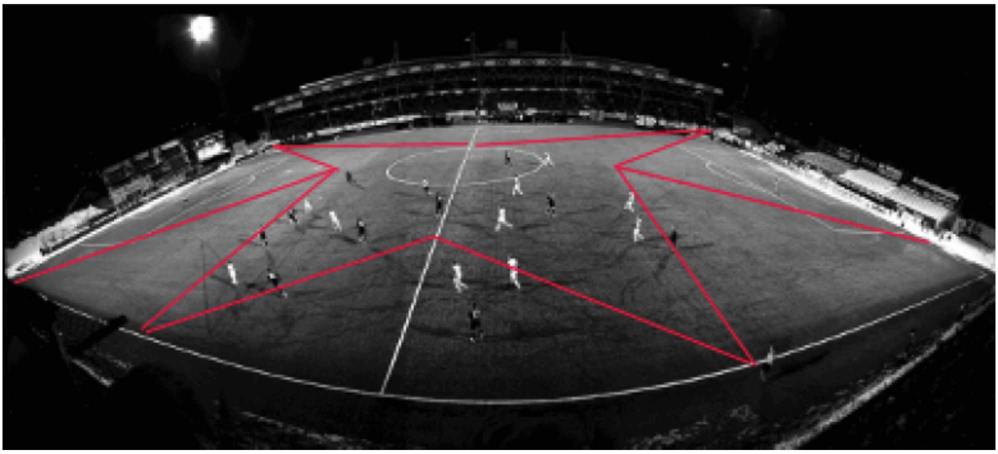

Fig. 3. Toggle zoom assignment.

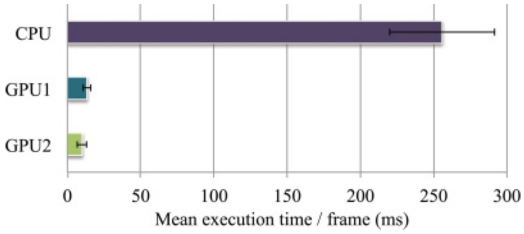

Fig. 4. Total execution time per frame in milliseconds.

center of the field as we move towards the goals. λ_0 is the zero order offset. All these parameters are empirically selected.

4.2.2. Toggle Zoom. This mode was developed to imitate the immediate switch in zoom levels. We picked a rather simple model for creating this effect. The panorama is partitioned into several zones and a focal length is assigned per zone. The zones can be seen in Figure 3.

5. EXPERIMENTAL RESULTS

To make an objective evaluation of the system, we have performed different sets of tests. The first set of tests evaluates the real-time properties of the virtual view generation. The rest of the experiments evaluate camera movements.

5.1. Execution Overhead

We have earlier proved that the given system can provide panorama video in real-time [Gaddam et al. 2014b], and in the context of generating a virtual view, we have tested three different implementations: 1) a CPU version just looping through all the pixels per frame; 2) a straight forward GPU port; and 3) an optimized GPU implementation where the system renders OpenGL textures written by an NVidia CUDA kernel directly from the GPU to the screen. The total per frame execution times for a virtual camera with full HD resolution on an Intel i7-2600 CPU with an Nvidia GeForce GTX 460 GPU is shown in Figure 4. Both GPU versions easily reach the

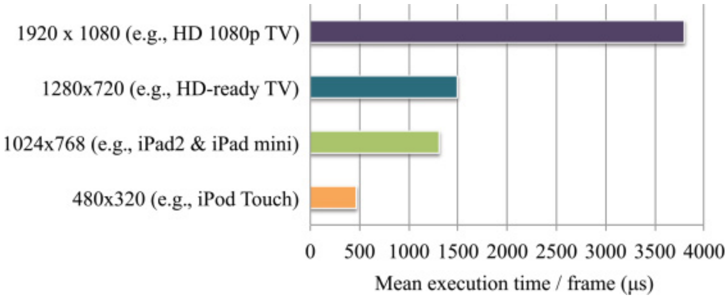

Fig. 5. Core execution times for various resolutions.

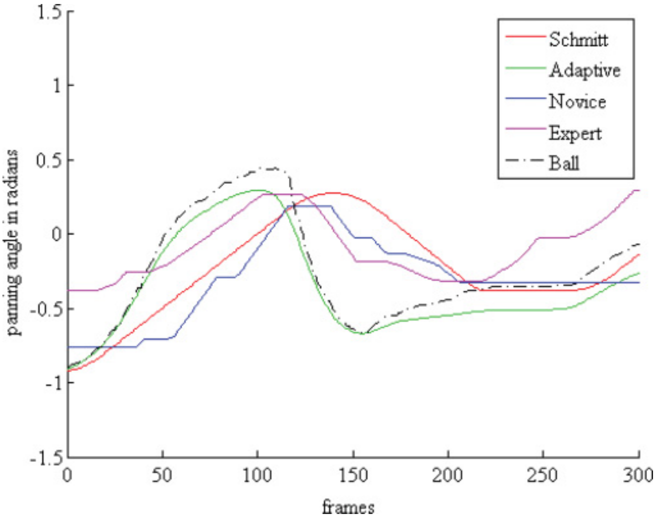

Fig. 6. Schmitt trigger and adaptive trigger plots for 300 frame segment along with the plots from human operated camera.

real-time requirements of 25 frames per second with an average of 13 and 10 milliseconds, respectively.

Now, not all receivers require a full HD video. In this respect, we have also evaluated the impact of interpolation and ray intersection costs depending on the size, that is, the only parts that varies with the output resolution. The results are shown in Figure 5. In short, the size of the output is negligible.

5.2. Pan/Tilt Models

The execution times for the Schmitt trigger and adaptive trigger are around 2μs and 30μs, respectively. Even though they differ by several orders, the absolute values are still negligible. Figure 6 provides a 300 frames segment for camera movements and ball position, where we see the pan/tilt angle (in radians) of the virtual view generated by both machine and human operations. In other words, if the curves are close, they capture more or less the same view. The causal nature of the Schmitt trigger and the human operators can be observed in the figure owing to the fact that, they get the ball position as it happens. On the other hand, the adaptive model has access to a small future data.

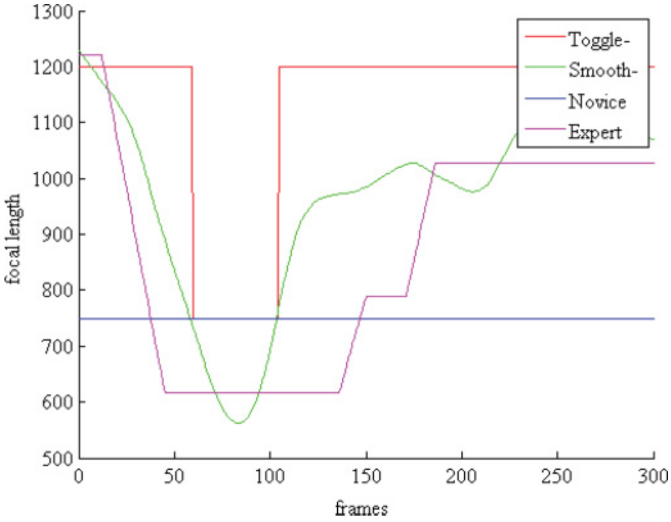

Fig. 7. Smooth zoom and toggle zoom plots along with the plots from human operated camera for 300 frame segment.

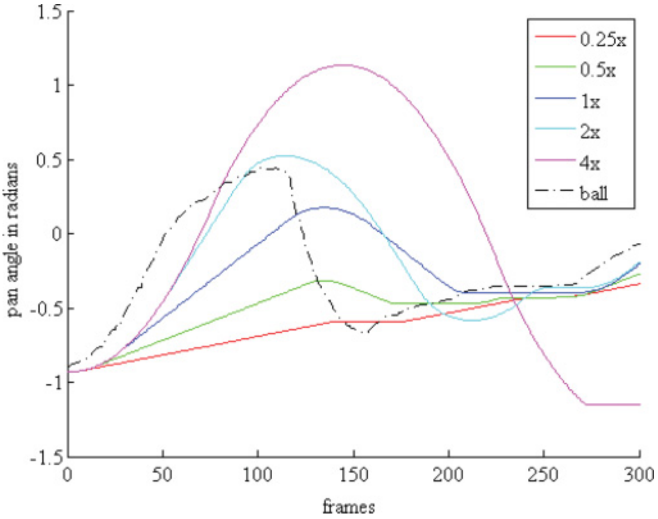

Fig. 8. The calculated trajectories for various acceleration values in the Schmitt trigger case.

5.3. Zoom Models

Since the calculation of zoom is a closed form expression over the current viewing position, the execution time is really low. Figure 7 provides plots from the different zoom models and the human operators over a 300 frames segment. Since the position is dependent on the pan/tilt model chosen, both curves are calculated using the adaptive trigger. It can be observed that the machine generated zoom curves show noticeable similarity to the expert controlled camera, irrespective of the simplicity in the models.

5.4. Schmitt Trigger—Analysis

There are three control parameters in the Schmitt trigger case. The acceleration, maximum velocity and the stopping-acceleration. Figure 8 displays the curves angle curves

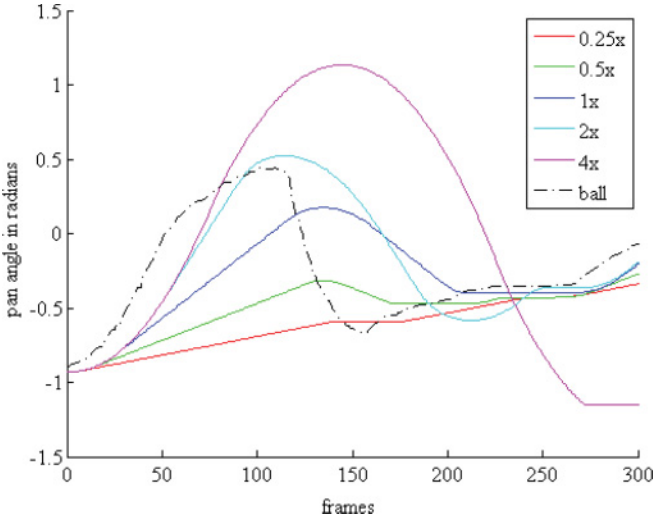

Fig. 9. The trajectories for various max velocities in the Schmitt trigger case.

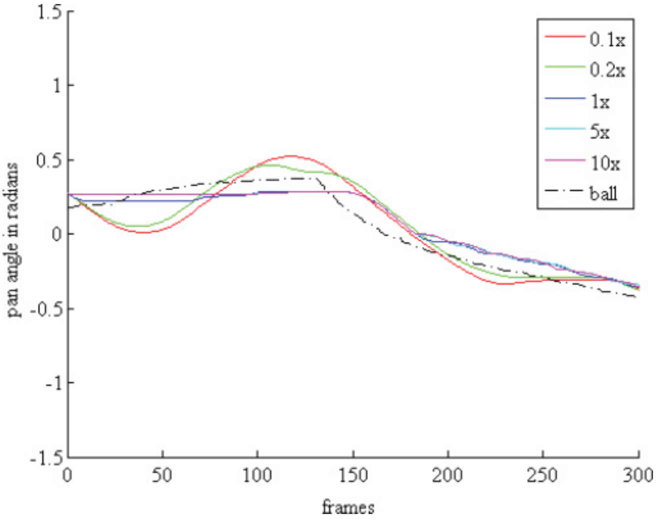

Fig. 10. Effect of varying stop-acceleration on the trajectories in Schmitt trigger case.

for different accelerations over 300 frames. It can be observed that higher acceleration tends to get the camera center closer to ball position quickly, but a problem is that it also introduces uneasiness in watching.

Figure 9 demonstrates the effect of varying the maximum velocity over 300 frames. When the ball moves really quickly, the curves in the plot show that the higher the maximum velocity, the closer they get to the slope required. However, this creates an undesired effect of overshooting irrespective of the quick deceleration.

Moreover, Figure 10 demonstrates the effect of the stop acceleration on the virtual camera movement. The trade-off here is between an appearance of a mechanical stop to a swinging effect. Both the velocity and acceleration effects can be seen in the plots.

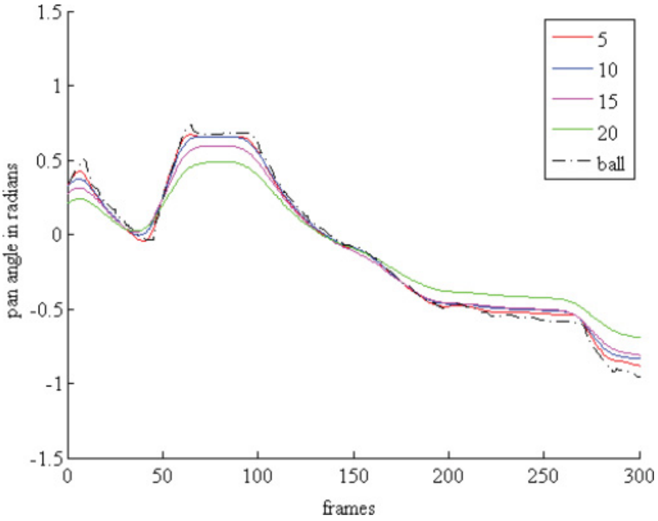

Fig. 11. Effect of window size selected for smoothing on the trajectories using the adaptive trigger.

5.5. Adaptive Trigger—Analysis

In the adaptive trigger, we have two control parameters. One is the window size and the other is thresholding for clipping. The thresholding for clipping only eliminates small jerky movements, so it is empirically chosen and it's plots do not provide great variation. Figure 11 demonstrates the effect on window size on the panning variable. The window size is varied between 5, 10, 15, and 20 frames. A scene of 300 frames where there are enough changes in the ball direction is picked. The exact field of view depends on the current focal length. However, as a rule of thumb, anything inside 0.2–0.5 radians from the center of the virtual camera can be assumed to be inside the field of view.

6. USER STUDIES

In the development of a user-centered system like ours, subjective feedback is essential to select the approaches giving the best quality of experience (QoE). Several experimental approaches [ITU-T 1998; ITU-R 2002] have been adapted and extended by researchers in the field of multimedia. Such QoE studies aim to assess, for example, behavioural responses to different aspects of multimedia systems [Wu et al. 2009], and particular attention has been devoted to the perception of video quality and the detection of visual artifacts [Farias et al. 2007; Goldmann et al. 2010; Ni et al. 2011]. To perform our assessment experiments, we have taken advantage of the flexibility of online tests which lately have become common [Chen et al. 2009]. Furthermore, since the user experience with the system is dependent on many factors outside video quality, we decided to introduce a pairwise comparison test [Lee et al. 2012; Ni et al. 2011] to contrast the different combinations of camera movements, two by two. When asked to select one of two versions of the same sequence, participants are presented with a task that is comparatively simpler than subjective ratings of sequences. Seeing how pairwise comparisons only require decisions on one's preference, this test is a good alternative when exposing participants to unfamiliar stimuli and situations [Lee et al. 2012].

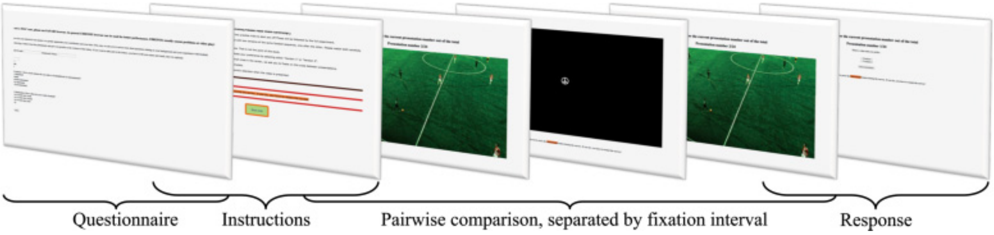

Fig. 12. Visual outline of the steps presented in the user study. Participants started with the questionnaire and instructions, before moving on to the soccer sequences. These were introduced by two practice trials, followed by the full study. Each pairwise comparison was separated by a 2-second fixation interval and terminated in a response session.

6.1. Study 1: Camera Controls

Because the system aims to provide users with the best possible experience, we need user feedback in order to establish the most preferable parameters for camera movements and zooming. We therefore conducted a user study to compare center trigger and adaptive camera movements, as well as toggle and smooth camera zooms.

6.1.1. Method. User preference for transient variables, such as camera movement and zoom, is deemed to be highly subjective and to depend on the presented sequence. To avoid subjective ratings that may vary more between presentations than between our experimental variables, we decided to use pairwise comparisons, as recommended by the ITU [ITU-T 1998]. Hence, each sequence was presented twice in a row, with only our variables of interest changing between presentations as shown in Figure 12.

Participants. A total of 49 users, 42 men and 7 women, participated in the first study. They were aged between 20 and 40 years, with an average of 27 years. Participants were presented with the opportunity to enter a lottery for a chance to win a small prize.

Stimuli and Procedure. All soccer sequences[2] were derived from the same international league match, recorded in 2013. While the ITU [ITU-T 1998] recommends a duration of approximately 10 seconds for pairwise comparisons of video presentations, we placed higher priority in ensuring that the soccer sequences contained more than one example of pan, zoom and tilt movements. Due to this, we extended the set sequence duration to 15 seconds. Automated camera movements were implemented subsequently, making sure that each movement and zoom contrast was presented four times. Each soccer sequence was therefore presented twice, separated by a two-second interval showing a fixation point on a black background. Stimuli contrasts were paired up so that either the camera movement or the camera zoom approach differed between the first and the second presentation. Although each paired contrast was presented four times, new soccer sequences were included for every pairwise comparison. Thus, participants watched 16 unique sequences, selected as the most suitable excerpts from the entire soccer match.

As stated before, we conducted the study using an online web-form so participants could complete it at their convenience. The paired video presentations were grouped in two stimuli blocks, with every contrast repeated twice within a block. Stimuli were counterbalanced with reverse-order for half of the contrasts, before they were randomised within each block. We created four randomised versions of the study, so that

[2] For a visual appreciation of the different approaches, two video sequences are included as examples for each of the different automatic and manual camera modes. These are attached with the submission.

Table I.
Parametric and non-parametric statistics for the number of times a stimulus combination was preferred over its contrasts, averaged across participants and sorted according to the Friedman rank score. Wilcoxon signed-rank test indicates statistically significant differences between stimuli, these are reported in relation to the lower ranked stimulus (the row above). Nonsignificant contrasts are labelled *ns*, while non-applicable comparisons are marked with a hyphen.

Results from Study 1								
Stimulus	**Parametric statistics**		**Percentiles**			**Friedman**	**Wilcoxon**	**Effect**
combination	**Mean**	**Std. dev.**	**25th**	**50th**	**75th**	**rank score**	**signed-rank test**	**size**
Schmitt/Toggle	1.77	1.40	1	2	3	1.32	–	–
Schmitt/Smooth	4.00	1.25	3	4	5	2.47	<.001	−0.51
Adaptive/Toggle	4.36	1.10	3	4	5	2.74	ns	−0.14
Adaptive/Smooth	5.87	1.56	5	6	7	3.47	<.001	−0.39
Results from Study 2								
Stimulus	**Parametric statistics**		**Percentiles**			**Friedman**	**Wilcoxon**	**Effect**
combination	**Mean**	**Std. dev.**	**25th**	**50th**	**75th**	**rank score**	**signed-rank test**	**size**
Novice	2.60	2.06	1	2	4	1.31	–	–
Expert	5.37	1.68	4	5	7	2.24	<.001	−0.52
Adaptive/Toggle	7.43	1.67	6	7	8	2.99	<.001	−0.41
Adaptive/Smooth	8.60	2.09	7	9	10	3.46	<.021	−0.28

the random order varied between participant groups. In order to control whether subjective preferences depended on soccer viewing experience, we introduced the study with two questions to assess soccer interest and dedication; we also collected details on age and gender. Participants received no information on the camera implementations, instead they received instructions to select the version they preferred. Following the questionnaire and instructions, we included two practice trials to get participants acquainted with the task, these were succeeded by the 16 pairwise comparisons.

6.1.2. Results. With every contrast repeated four times, the preference scores for the different conditions were added up for every participant. This resulted in individual counts the four combinations of camera movements and camera zooms, ranging from 0 to 4. In order to identify and weed out outlying preference counts, we also calculated the difference in scores between paired stimuli. This resulted in four mean differences, and we used the average of these to identify any scores that fell more than two standard deviations from the mean. Accordingly, we identified and excluded data from two participants, whose mean difference scores of zero indicated that they were unable to distinguish between stimuli. For the main analysis, we collapsed preference scores across stimulus combinations to obtain the overall number of times each camera mode was preferred by an individual. With two contrasts repeated four times for every camera mode, the highest possible preference count comes to 8. A Friedman rank test was used to analyse the preference counts from the remaining 47 participants, revealing a significant effect of our camera implementations ($\chi^2(3) = 72.73$). To further explore the difference between stimulus combinations, we also ran three Wilcoxon signed-rank tests and calculated effect sizes from these. In addition to the non-parametric tests, parametric means and standard deviations are included to better highlight the distribution of scores. Results from the analyses are presented in Table I. Furthermore, we also explored the individual contrasts with a Friedman rank test, again revealing a significant overall effect ($\chi^2(7) = 162.33$). These results are illustrated in Figure 13, listed according to their Friedman rank scores.

From the collapsed preference counts and the ranking scores presented in the first part of Table I, the adaptive trigger movement combined with the smooth focal zoom

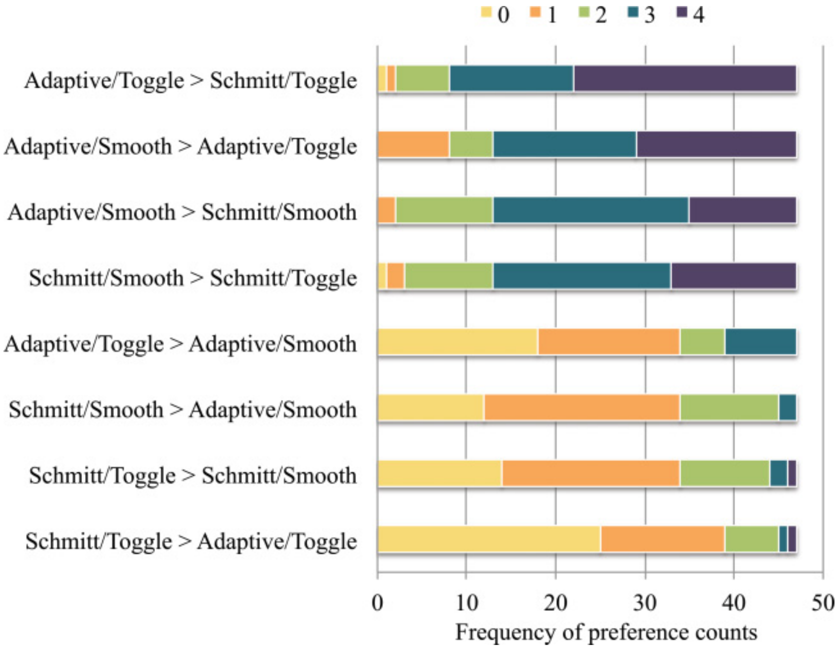

Fig. 13. Frequency distribution portraying the number of times one stimulus was preferred over its contrast, accumulated across users. For example, in the first line, we see that 25 persons have preferred the Adaptive/Toggle over Schmitt/Toggle in all four repetitions. The maximum count of 4 corresponds to the number of repetitions for each pair of videos. Stimulus contrasts are sorted according to Friedman rank scores and plotted symmetrically.

emerges as the preferred camera implementation. Although not significantly different from the third rank, the adaptive trigger remained the preferred choice over the Schmitt trigger, ranking second when combined with the toggle focal zoom. These trends are also evident when looking at the ranked individual contrasts in Figure 13. The adaptive trigger movement is preferred over Schmitt alternative for the vast majority of presentations, just as the smooth focal is the predominantly preferred zoom option over the toggle focal. In short, the opinions of 47 users clearly demonstrate the preference for the adaptive trigger and smooth focal camera implementation.

6.2. Study 2: Man vs. Machine

Following the results from Study 1, we established that users prefer the camera movement combination with adaptive trigger pan and smooth focal zoom. However, an important challenge for such an automated system is to provide a viewing experience that can compete with a soccer match filmed by a manually operated camera. Hence, the second user study compares user preferences for the two highest ranked automated camera implementations with that of two human operators.

6.2.1. Method. The second user study was conducted as a pairwise comparisons test, with the same setup used for Study 1.

Participants. With 14 females and 23 males, we collected data from 37 participants, none of whom had taken part in Study 1. Their ages spanned from 21 to 71 years, with an average of 29 years. Every participant was provided with the opportunity to sign up for a lottery that offered small prizes to be won.

Stimuli and Procedure. To compare automated camera movements with manual camera operations, we selected the two best-preferred stimulus combinations from Study 1. In so doing, we re-used half of the stimuli from the first user study and compared these to sequences with recorded camera movements. To record the camera movements, we invited an expert and a novice camera operator to watch the same soccer match. The expert was an experienced camera operator from a Scandinavian broadcaster, whereas the novice had experience with camera-view operations within games. After receiving instructions on how to move and zoom with the virtual camera using a joystick, the operators embarked upon the task of following the match by keeping the ball and action in focus. From their recordings, we selected 20 expert and 20 novice 15-second excerpts to contrast with the automated sequences. For further verification of the preference ratings from Study 1, we also contrasted the automated sequences with each other. Moreover, we contrasted the expert and novice recordings to see whether preferences differed between the two.

Study 2 proceeded in the same manner as Study 1, described in Section 6.1.1. The only procedural distinction between the two studies is the inclusion of more stimuli, resulting in 24 pairwise comparisons.

6.2.2. Results. Response data from Study 2 were restructured and analysed the same way as described for Study 1 in Section 6.1.2, again with 2 outliers detected and excluded. With the Friedman rank test indicating significant differences between the collapsed preference counts ($\chi^2(3) = 56.73$), we again followed up with Wilcoxon signed-rank tests. Results from these analyses are included in Table I. A second Friedman rank test revealed significant differences also between the individual contrasts ($\chi^2(11) = 177.15$), the ranked preference counts for these are portrayed in Figure 14.

First and foremost, the results from Study 2 reveal that users clearly prefer automated over manual camera movements. Of course, the quality of manual controls is only as good as the operator. We considered this possible limitation and took precautions by including two camera operators, one expert and one novice. The higher ranking of the expert over the novice operator exemplifies the importance of the camera man's expertise. Despite our precautions, we cannot ascertain that users will prefer the automatic camera operations over any camera operator. However, considering the significant differences and the magnitudes of effect sizes for the presented conditions, the results show that our system outperforms the two human operators. Specifically, a consistent trend can be observed for both the collapsed preference counts (Table I) and the individual contrasts (Figure 14), where the automated camera movements are chosen over the manual operations in the majority of presentations. Furthermore, the higher rank for the adaptive/smooth over the adaptive/toggle combination reflects the results from Study 1.

7. DISCUSSION

The motivation behind designing and developing such a system lies in providing an interaction to the user. In cases where manual control of the virtual camera is desired, the system simplifies significantly. On the other hand, a viewer following a game might be interested in interaction but at a higher level. The viewer might place a request to the client to follow the ball/a single player or a collection of players. In such a case, the client has to provide an aesthetically pleasing virtual camera based on the position data from the ball and players. Even a coach is greatly advantaged by such a system, he/she can instantly request multiple virtual cameras focussing on different features. For example, one for the ball, one for a recently injured player, one for a recently exchanged player and one for the defense. So, building the entire system and a subjective evaluation of the results proved to be mandatory.

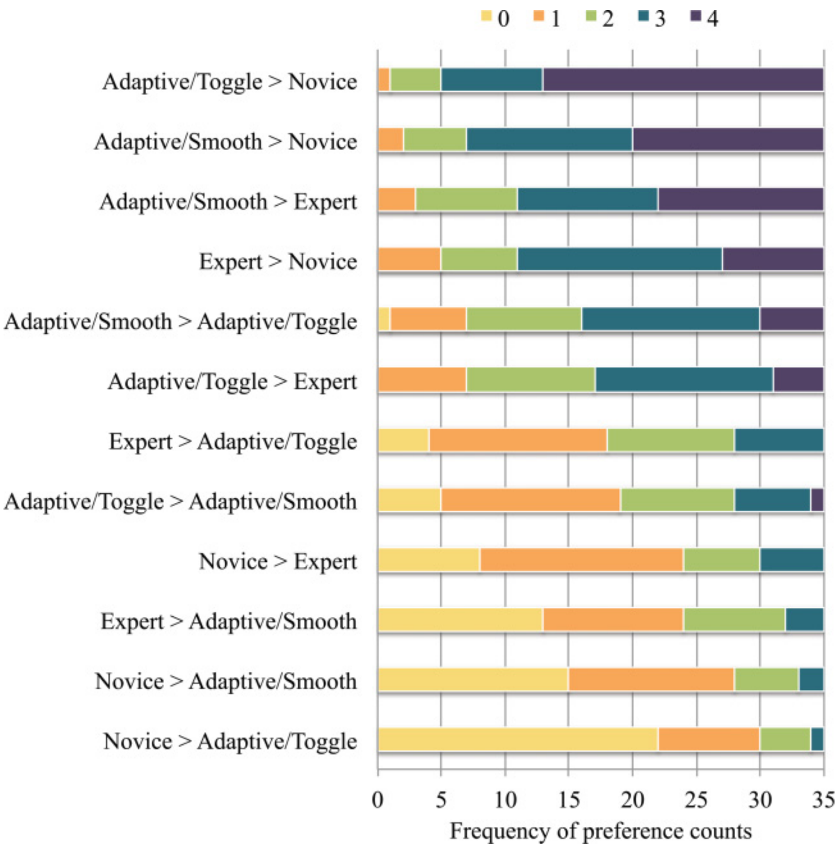

Fig. 14. Frequency distribution portraying the number of times one stimulus was preferred over its contrast, accumulated across users. The maximum count of 4 corresponds to the number of repetitions. Stimulus contrasts are sorted according to Friedman rank scores and plotted symmetrically.

In the two user studies, we have explored and analysed user preferences for automated and manual camera movements. The first study established that the average user prefers the adaptive trigger movement over the Schmitt trigger and the smooth focal zoom over the toggle; implications of these findings are discussed further on. From the second user study, we found that the average user maintains the same preference for the adaptive trigger and the smooth focal zoom when compared to a human-operated camera. While this finding is specific to the current context and may not reflect the performance of all camera operators, the subjective preference for automated camera movements suggests a positive user experience with our system. Overall, the presented results are promising for the future acceptance and use of our system.

The user preferences between the toggle and smooth zoom is slightly ambiguous. From the user study, it is clear that the smooth zoom is preferred, but toggle zoom provides the advantage to switching to an overview immediately. This when combined with smooth zoom for smaller ball changes can provide a nice aesthetic, yet functional camera motion that can keep the ball in field of view. Moreover, the zoom model currently is based only on the position of the ball on the panorama. This can be significantly improved by incorporating game context into the model. Some of the things can be velocity of the ball, player arrangement and special events (penalty, corner or throw-in).

Furthermore, it must be noted that this study focuses on one of the several points from where the action is captured on the soccer field. When it comes to capturing from one point in live, the camera man has little freedom in the grammar of the video. In an actual broadcast, the producer mixes several streams together and this is where the grammar come into place.

In future, we are aiming to improve several components of the system. We are currently working on capturing High Dynamic Range (HDR) panoramic videos to handle the loss of details in shadows on sunny days. We are also investing our energy into developing the client on a mobile platform, which has its own challenges concerning the bandwidth and power consumption.

Moreover current day's visual tracking algorithms' recall is not practically applicable to real-life scenarios. Owing to this, we still have a large manual component when it comes to estimating the ball position. We are currently exploring algorithms based on multi-sensor data to track the ball with a high recall rate. When we track the ball successfully, we will be able to provide a complete system functional in real time. However, we do have an accurate tracking of the player positions, meaning that the system easily can follow a single player or a group of players.

8. CONCLUSION

In our research, we have shown that a single camera-array generated panorama video can support an arbitrary number of virtual views, which are generated locally on the client device. In many scenarios, users will want to control and interact with their own virtual camera, whereas other situations require higher levels of abstraction. In order to generate video streams that incorporate automated camera movements while satisfying user expectations, we have explored machine-controlled camera modes versus human camera operators in two separate user studies. In the first, we explored automatic movement approaches and established that the best-preferred mode combines smooth focal zoom with adaptive trigger movements. The second study compared machine and human generated camera movements, with results that promise well for future acceptance of a machine controlled cameraman. However, It must be noted that we are not claiming a system that is capable of exceeding a human operator. The research outcomes here do not guarantee an assertion that a machine can beat a human. The study merely points at the one of the several possible positive futures in the current context and scenario. There are several many variables and a long way to generalize and extend this to the entire broadcasting paradigm. Our ongoing work include both improved object tracking and further parallelization; most importantly though, we aim to further improve the automated camera movements.

REFERENCES

Adel Ahmed and Peter Eades. 2005. Automatic camera path generation for graph navigation in 3D. In *Proceedings of the Asia-Pacific Symposium on Information Visualisation*. 27–32. http://dl.acm.org/citation.cfm?id=1082315.1082320

Y. Ariki, S. Kubota, and M. Kumano. 2006. Automatic production system of soccer sports video by digital camera work based on situation recognition. In *Proceedings of the IEEE International Symposium on Multimedia*. 851–860. DOI:http://dx.doi.org/10.1109/ISM.2006.37

Peter Carr and Richard Hartley. 2009. Portable multi-megapixel camera with real-time recording and playback. In *Proceedings of the Conference on Digital Image Computing: Techniques and Applications*. 74–80. DOI:http://dx.doi.org/10.1109/DICTA.2009.62

Peter Carr, Michael Mistry, and Iain Matthews. 2013. Hybrid robotic/virtual pan-tilt-zom cameras for autonomous event recording. In *Proceedings of the ACM Multimedia Conference*. 193–202.

Joel Carranza, Christian Theobalt, Marcus A. Magnor, and Hans-Peter Seidel. 2003. Free viewpoint video of human actors. *ACM Trans. Graph.* 22, 3, 569–577. DOI:http://dx.doi.org/10.1145/882262.882309

Fan Chen and Christophe De Vleeschouwer. 2010. Personalized production of basketball videos from multi-sensored data under limited display resolution. *Computer Vision Image Understanding* 114, 6, 667–680. DOI:http://dx.doi.org/10.1016/j.cviu.2010.01.005

Kuan-Ta Chen, Chen-Chi Wu, Yu-Chun Chang, and Chin-Laung Lei. 2009. A crowd-sourceable QoE evaluation framework for multimedia content. In *Proceedings of the ACM Multimedia Conference*. 491–500. DOI:http://dx.doi.org/10.1145/1631272.1631339

Shenchang Eric Chen. 1995. QuickTime VR: An image-based approach to virtual environment navigation. In *Proceedings of the ACM SIGGRAPH International Conference on Computer Graphics and Interactive Techniques*. 29–38. DOI:http://dx.doi.org/10.1145/218380.218395

Marc Christie, Rumesh Machap, Jean-Marie Normand, Patrick Olivier, and Jonathan Pickering. 2005. Virtual camera planning: A survey. In *Smart Graphics*, Lecture Notes in Computer Science, vol. 3638, 40–52. DOI:http://dx.doi.org/10.1007/11536482 4

A Dearden, Y Demiris, and O Grau. 2007. Learning models of camera control for imitation in football matches. In *Proceedings of the Artificial and Ambient Intelligence Symposium*. 227–231.

Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. 1996. Modeling and Rendering Architecture from Photographs: A hybrid geometry- and image-based approach. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'96)*. ACM, New York, 11–20. DOI:http://dx.doi.org/10.1145/237170.237191

Mylène C. Q. Farias, John M. Foley, and Sanjit K. Mitra. 2007. Detectability and annoyance of synthetic blocky, blurry, noisy, and ringing artifacts. *IEEE Trans. Signal Process.* 55, 6, 2954–2964. DOI:http://dx.doi.org/10.1109/TSP.2007.893963

Christoph Fehn, Christian Weissig, Ingo Feldmann, Markus Muller, Peter Eisert, Peter Kauff, and Hans Bloss. 2006. Creation of high-resolution video panoramas of sport events. In *Proceedings of the IEEE International Symposium on Multimedia*. 291–298. DOI:http://dx.doi.org/10.1109/ISM.2006.55

Eric Foote, Peter Carr, Patrick Lucey, Yaser Sheikh, and Iain Matthews. 2013. One-man-band: A touch screen interface for producing live multi-camera sports broadcasts. In *Proceedings of the ACM Multimedia Conference*. 163–172. DOI:http://dx.doi.org/10.1145/2502081.2502092

Vamsidhar Reddy Gaddam, Carsten Griwodz, and Pål Halvorsen. 2014a. Automatic exposure for panoramic systems in uncontrolled lighting conditions: a football stadium case study. In *Proceedings of SPIE: The Engineering Reality of Virtual Reality*. 90120C–90120C–9. DOI:http://dx.doi.org/10.1117/12.2040145

Vamsidhar Reddy Gaddam, Ragnar Langseth, Sigurd Ljødal, Pierre Gurdjos, Vincent Charvillat, Carsten Griwodz, and Pål Halvorsen. 2014b. Interactive Zoom and Panning from Live Panoramic Video. In *Proceedings of the ACM International Workshop on Network and Operating Systems Support for Digital Audio and Video*. Article 19. DOI:http://dx.doi.org/10.1145/2578260.2578264

Lutz Goldmann, Francesca De Simone, Frederic Dufaux, Touradj Ebrahimi, Rudolf Tanner, and Mauro Lattuada. 2010. Impact of video transcoding artifacts on the subjective quality. In *Proceedings of the International Workshop on Quality of Multimedia Experience*. 52–57.

Patrik Goorts, Steven Maesen, Maarten Dumont, Sammy Rogmans, and Philippe Bekaert. 2014. Free viewpoint video for soccer using histogram-based validity maps in plane sweeping. In *Proceedings of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. 378–386.

O. Grau, T. Pullen, and G. A. Thomas. 2004. A combined studio production system for 3-D capturing of live action and immersive actor feedback. *IEEE Trans. Circuits Syst. Video Technol.* 14, 3, 370–380. DOI:http://dx.doi.org/10.1109/TCSVT.2004.823397

O. Grau, G. A. Thomas, A. Hilton, J. Kilner, and J. Starck. 2007. A robust free-viewpoint video system for sport scenes. In *Proceedings of the 3DTV Conference*. 1–4. DOI:http://dx.doi.org/10.1109/3DTV.2007.4379384

Pål Halvorsen, Simen Sægrov, Asgeir Mortensen, David K. C. Kristensen, Alexander Eichhorn, Magnus Stenhaug, Stian Dahl, Håkon Kvale Stensland, Vamsidhar Reddy Gaddam, Carsten Griwodz, and Dag Johansen. 2013. BAGADUS: An Integrated system for arena sports analytics – A soccer case study. In *Proceedings of the ACM Multimedia Conference*. 48–59.

S. Hutchinson, G. D. Hager, and P. I. Corke. 1996. A tutorial on visual servo control. *IEEE Trans. Rob. Automation* 12, 5, 651–670. DOI:http://dx.doi.org/10.1109/70.538972

ITU-R. 2002. BT.500-11. Methodology for the subjective assessment of the quality of television pictures. https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-11-200206-SIIPDF-E.pdf.

ITU-T. 1998. P.911. Subjective audiovisual quality assessment methods for multimedia applications. https://www.itu.int/rec/T-REC-P.911-199812-1/en.

Michael Jenkin, James Elder, and Greg Pintilie. 1998. Loosely-coupled telepresence through the panoramic image server. In *Vision Interface: Real World Applications of Computer Vision*.

R. Kaiser, M. Thaler, A. Kriechbaum, H. Fassold, W. Bailer, and J. Rosner. 2011. Real-time person tracking in high-resolution panoramic video for automated broadcast production. In *Proceedings of the European Conference on Visual Media Production*. 21–29. DOI:http://dx.doi.org/10.1109/CVMP.2011.9

Takeo Kanade, Peter Rander, and P. J. Narayanan. 1997. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE MultiMedia* 4, 1, 34–47. DOI:http://dx.doi.org/10.1109/93.580394

Jong-Seok Lee, Lutz Goldmann, and Touradj Ebrahimi. 2012. Paired comparison-based subjective quality assessment of stereoscopic images. *Multimedia Tools Appl.* 67, 1, 31–48. DOI:http://dx.doi.org/10.1007/s11042-012-1011-6

Christian Lipski, Christian Linz, Kai Berger, and Marcus Magnor. 2009. Virtual video camera: Image-based viewpoint navigation through space and time. In *Proceedings of the ACM SIGGRAPH International Conference on Computer Graphics and Interactive Techniques*. Article 93. DOI:http://dx.doi.org/10.1145/1599301.1599394

Aditya Mavlankar and Bernd Girod. 2010. Video streaming with interactive pan/tilt/zoom. In *High-Quality Visual Experience*, Marta Mrak, Mislav Grgic, and Murat Kunt (Eds.), 431–455. DOI:http://dx.doi.org/10.1007/978-3-642-12802-8 19

Pengpeng Ni, Ragnhild Eg, Alexander Eichhorn, Carsten Griwodz, and Pål Halvorsen. 2011. Flicker effects in adaptive video streaming to handheld devices. In *Proceedings of the ACM Multimedia Conference*. 463–472.

N. Papadakis, A. Baeza, I. Rius, X. Armangue, A. Bugeau, O. D'Hondt, P. Gargallo, V. Caselles, and S. Sagas. 2010. Virtual camera synthesis for soccer game replays. In *Proceedings of the Conference on Visual Media Production*. 97–106. DOI:http://dx.doi.org/10.1109/CVMP.2010.20

Jinchang Ren, Ming Xu, James Orwell, and GraemeA. Jones. 2010. Multi-camera video surveillance for real-time analysis and reconstruction of soccer games. *Machine Vision Appl.* 21, 6, 855–863. DOI:http://dx.doi.org/10.1007/s00138-009-0212-0

Xinding Sun, J. Foote, D. Kimber, and B. S. Manjunath. 2005. Region of interest extraction and virtual camera control based on panoramic video capturing. *IEEE Trans. Multimedia* 7, 5, 981–990. DOI:http://dx.doi.org/10.1109/TMM.2005.854388

Marius Tennøe, Espen Helgedagsrud, Mikkel Næss, Henrik Kjus Alstad, Håkon Kvale Stensland, Vamsidhar Reddy Gaddam, Dag Johansen, Carsten Griwodz, and Pål Halvorsen. 2013. Efficient implementation and processing of a real-time panorama video pipeline. In *Proceedings of the IEEE International Symposium on Multimedia*.

Jinjun Wang, Changsheng Xu, Engsiong Chng, Kongwah Wah, and Qi Tian. 2004. Automatic replay generation for soccer video broadcasting. In *Proceedings of the ACM Multimedia Conference*. 32–39. DOI:http://dx.doi.org/10.1145/1027527.1027535

Wanmin Wu, Ahsan Arefin, Raoul Rivas, Klara Nahrstedt, Renata M. Sheppard, and Zhenyu Yang. 2009. Quality of experience in distributed interactive multimedia environments: Toward a theoretical framework. In *Proceedings of the ACM Multimedia Conference*. 481–490.

M. Xu, J. Orwell, L. Lowey, and D. Thirde. 2005. Architecture and algorithms for tracking football players with multiple cameras. In *IEE Proc. Vision Image Signal Process*. 152, 2, 232–241. DOI:http://dx.doi.org/10.1049/ip-vis:20041257

Wei Xu and Jane Mulligan. 2013. Panoramic video stitching from commodity HDTV cameras. *Multimedia Systems* 19, 5, 407–426. DOI:http://dx.doi.org/10.1007/s00530-013-0316-2

T. Yokoi and H. Fujiyoshi. 2005. Virtual camerawork for generating lecture video from high resolution images. In *Proceedings of the IEEE International Conference on Multimedia and Expo*. DOI:http://dx.doi.org/10.1109/ICME.2005.1521532

Xinguo Yu, Changsheng Xu, Hon Wai Leong, Qi Tian, Qing Tang, and Kong Wah Wan. 2003. Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. In *Proceedings of the ACM Multimedia Conference*. 11–20. DOI:http://dx.doi.org/10.1145/957013.957018

Received July 2014; revised November 2014; accepted February 2015