

You've Got the Moves, We've Got the Motion – Understanding and Designing for Cinematographic Camera Motion Control

Axel Hoesl^(✉), Partrick Mörwald, Philipp Burgdorf, Elisabeth Dreßler,
and Andreas Butz

LMU Munich, Munich, Germany
{axel.hoesl, andreas.butz}@ifi.lmu.de,
{moerwal, burgdorf, dressler}@cip.ifi.lmu.de

Abstract. Moving a film camera aesthetically is complex, even for professionals. They commonly use mechanical tools which help them to control camera motion. In recent years, computer-controlled tools were developed, but their development is mostly technology-driven and often fails to thoroughly integrate the user perspective. In HCI, prototyping is an established way to collect early feedback and thereby integrate a user perspective early on. In filmmaking, there is a lack of prototyping platforms, mostly due to the small market and inherent technical complexity of tools. We therefore developed a prototyping platform in cooperation between experts in camera operation, mechanical engineering and computer science. It consists of a motion control system for sliding camera moves composed of affordable hardware and open source software, and it supports the wireless connection of various types of user interfaces via Bluetooth. In this combination, it allows the exploration of different interface and control strategies in-the-wild, as it is easy to transport and stable for use in the field. A prototype using our platform was used by professional filmmakers in real commercial assignments. We further report on its use in two studies ($N = 18$, $N = 12$) examining the effects of various degrees of automation (low and medium) on the sense and quality of control. Our results indicate no decrease in both dimensions.

Keywords: Cinematography · Camera motion · Motion control · Slider

1 Introduction

Mastering precise and aesthetic camera motion in filmmaking is central to camera operators. Performing it manually is hard and errors are likely. To reduce errors, task sharing and the use of support tools have been established. In one single camera move, usually three – sometimes even more – operators work together simultaneously in a choreography, performed *behind* the camera [21]. In this process, subtasks are delegated to human operators or to machines. Both

perform very well in different areas: Humans, for example, outperform machines in recognizing visual patterns and aesthetic judgment; in contrast, machines can move heavy weights smoothly, precisely and repeatedly [16]. Originally, many of such support tools were purely mechanic. With advances in microelectronics, however, they were extended by motors and controlled by microprocessors starting in the late 1970s¹. Fueled by further technological advances in the decades to follow, a multitude of novel tools was introduced and is now summoned under the label of *camera motion control systems*. Going beyond the pure mechanics, nowadays high-tech tools such as industrial robots [6] or drones [17] became part of the tool palette in camerawork.

We identified several challenges, that currently hinder research in this field: On the market, there are mainly *expensive tools without access* for connecting new user interface (UI) prototypes. *Expertise in multiple fields*, such as mechanical engineering, electronics, human-computer interaction (HCI) and computer science is necessary in order to build them. This makes it hard to quickly translate new ideas into prototypes. Furthermore, the research literature on *physical cinematographic camera motion* and its control is not very elaborate to the best of our knowledge. There is a lack of ethnography, studies on systems and interaction designs and their evaluation for user-centered research. In the field of *virtual camera motion* there is plenty of literature available, but its findings cannot simply be adapted to physical camera control to suit the needs of enthusiasts and professionals on location. Existing support tools are meant to be used in the physical world and also serve the artistic expression. The latter often involves a trial and error experience of unforeseen dynamic changes depending on how a situation unfolds. Therefore, their use and control is *hard to simulate* in a virtual environment [27]. In addition, operators want to delegate tasks, but also want to be in control of the recorded images [21]. Delegation and being in control, however, are often contradictory [28]. Therefore finding the right balance for different user groups is non-trivial. To be meaningful to operators, systems therefore need to *balance* user control with automation [21]. This might be achieved best by introducing *high level* controls, but further research on systems and interactions is necessary.

1.1 Contribution

We contribute an open source motion-controlled camera slider with independent control and powering units that is *inexpensive* and offers open *wireless* access. With this platform, various types of UIs balancing delegation with control can be prototyped and evaluated. It was tested by a professional cinematographer in five assignment shootings. We found strong indication for its *high quality in smooth motion* and *system stability*. We then used it to conduct two controlled user studies. In our first study, we examined the effects of motorized tools for camera motion on the *sense* and *quality* of control of the participants. In a subsequent

¹ The first major motorized and computer controlled system of this kind was the Dykstraflex [3], used for special effects in Star Wars Episode IV: A New Hope (1977).

study, we investigating the influence of *reviewing* results by the participants during the evaluation process of the recorded material. This helped us to validate the effects on the participants' sense of control.

2 Related Work

In the domain of *virtual* camera motion, the specifics of camera control were summarized by Bowman et al. [5] and Christie et al. [14] classifying approaches, analyzing requirements and revealing limitations. However, not all of the presented approaches can simply be translated to real world cinematography, as, for instance, the scene in hand concept [34]. Suitable approaches – especially for high level control – are often image-based [26] or constraint-based [13, 23]. Here, Through-The-Lens controls [12, 18, 23] for constraint-based camera positioning combined with 3D navigation techniques for direct control – i.e., lower level axis control – e.g., using multi-touch gestures on mobile devices as Move&Look [25], have the potential to correspond well to established mental models of users in cinematography and 3D navigation.

In *physical* camera motion, a survey on autonomous camera systems recently conducted by Chen and Carr [9] identified the core tasks and summarized twenty years of research-driven tool development and evaluation. Within this domain, research tools are found in multiple areas sometimes going beyond traditional cinematography. Autonomous pan-and-tilt cameras are used to autonomously record academic lectures. In the work of Hulens et al. [22], the Rule of Thirds² is borrowed from cinematography and integrated into the tracking. Zhang and colleagues [36] presented a tele-conferencing system incorporating video and context analysis. In their work, the camera is oriented and zoomed automatically in order to better guide the users' attention. If, for example, a presenter shows certain details on a whiteboard, this area is automatically zoomed in on. For automated sports broadcasting [8], Chen et al. [10] even mimic the operation style of human operators through machine learning as the automated operation style is often perceived as rather “robotic”.

The presented examples implement machine benefits, but hardly offer a *human-machine interplay* allowing operators to contribute. One of the few examples offering such an interplay is presented by Stanciu and colleagues [31]. Here, a crane with a camera mounted on one side automatically frames a user-selected target and adapts to the manual crane operation of a human operator on the opposite side³. This human-machine interplay is important for the operators as they want to actively express their personal view and therefore want to feel in control [21]. For interactive control, a futuristic vision of novel and natural forms of interaction was already presented with Starfire [32] in 1994. The video

² A guideline for image composition where the image frame is divided into three thirds, both horizontally and vertically. Important subjects, e.g., a speaker, are best placed at one of the intersections.

³ Stypekit [4] is a corporate implementation of the concept.

prototype showcased a camera crane that was controlled by a tablet used as a remote viewport. So far however, no implementation followed the concept⁴.

3 Prototype Development

Advanced tools, such as camera cranes [4], are complex to manage, since multiple degrees of freedom (DOF) need to be controlled in real time. This usually requires years of training, is expensive, complex to build and requires high efforts in transportation. For the quick translation of new interaction concepts, not all of the offered DOF are always necessary in order to evaluate alternative designs on a conceptual level. As pointed out by Nielsen [29], filmmakers do not always apply all possible movements and exploit all DOFs. They rather carefully choose DOFs and movements as stylistic devices depending on the content of the scene. Often shots that are outside the “regular” human visual experience need to be motivated in particular by the content and are thus used less often. Our prototype supports some go-to types of shots that can often be seen in diverse contexts such as advertising, image, short, feature and documentary films.

3.1 Collecting Expert Requirements

Before we developed our prototype, we interviewed professionals in camera motion ($N=3$) to determine user roles when using camera sliders in their work.

User Roles. We handed out an online questionnaire to camera operators. Focusing on qualitative statements, we distinguished two usage contexts: *solo operation*, where one operator is out alone on location and thus needs to control everything that is usually delegated to multiple assistants. We also found mentions of the classic collaboration with *task delegation*, where tasks are shared and delegated to separate operators. This is often used on film sets.

Hardware and User Interface Requirements. An additional professional operator was interviewed about technical hardware and user interface requirements. He was experienced in camera operation and in consulting major producers of cinematographic equipment. From a 45 min semi-structured interview we derived the following requirements (Table 1).

3.2 Implementation

Based on the identified scenarios and requirements, we built a setup focusing on *solo operation* with *task delegation* to an assisting system. Together with a mechanical engineer, we determined the necessary motor torque and acceleration for actuating the payload of 20 kg horizontally and 6 kg vertically. An overview of

⁴ The Freefly MIMIC [1], however, reminds us of the original idea.

Table 1. Requirements as collected in the semi-structured interview with a professional

Hardware requirements	User interface requirements
Stable for on-set use	Wireless remote control
Smooth motion with stable images	Speed is controlled manually
Runs at constant speed	Programming of moves
Bounces between ends repeatedly	Programming of ramps
Stops before it hits an end	Moves can be repeated precisely
Offers time-lapse recordings	Moves can be saved, loaded and edited
Payload of 20 kg (hor.) and 6 kg (vert.)	Displays the camera stream live

the main units is presented in Fig. 1. The components, wiring diagrams, source codes of the control software, plans of the 3D-printed parts, documentation of the wireless control protocol (using Bluetooth Low Energy) and sample footage are provided electronically [20] in further detail. The final implementation is presented in the following section in Fig. 2.

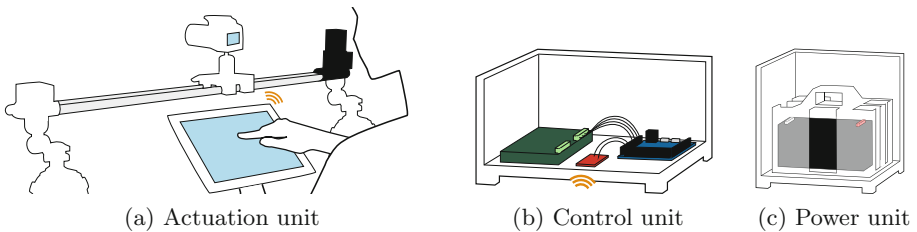


Fig. 1. In 1a an image-based [26] control scenario is illustrated. Instructions are sent to the control unit (1b) driving the motor (1a, right). Here, the signals are received by a Bluetooth module (1b, center) and processed by an Arduino (1b, right) controlling the motor driver (1b, left). The power unit (1c) supplies the system with electricity.

4 Field Evaluation

To determine whether the hardware requirements were met, we asked a professional camera operator to evaluate our setup during five assigned shootings. We made sure that he did not know the identified requirements to avoid a bias. The features concerned with speed changes, bouncing and safety stops were essential and thus already tested during the development of the hardware and firmware.

The requirements of *system stability* and *smoothness of motion* hence remained to be verified. The setup was used in five on-assignment shootings for exhibition films in a modern art museum (Fig. 2). At these recording sessions the system worked in a stable way. Horizontal shots were recorded at ground level and at waist height. Also, diagonal shots from waist height to ground level

and vertical shots at a height beyond two meters were recorded. The cameras were moved with constant speed and bounced between both ends for longer periods of time. The limit switches were used for each calibration and prevented the slider from hitting an end. Moves and ramps were programmed remotely and wirelessly by a second trained person using a laptop connected to the system via Bluetooth (Fig. 2).



Fig. 2. Field evaluation of the prototype in different scenarios with a professional camera operator during an assignment for exhibition films of a modern art museum.

4.1 Results and Discussion

Because of its reputation, the museum set high standards for the aesthetic and technical quality of its representation. The material recorded with our system eventually appeared in six exhibition films and was approved and published by the museum. We take this as an indicator for our system's capability to produce acceptable results. Only one recording session was planned in the beginning. The additional five sessions were initiated by the operator only after the results of the first were screened. In a summarizing debriefing the operator was generally satisfied with the systems stability, but also revealed some issues he found.

His main concern was that *saving time* is crucial. Therefore *shortcuts* should be provided. Often he would start with a slide from one end to the other at medium speed and then adjust positioning or speed depending on what he saw on the camera display. We therefore added this feature to the UI requirement list. In our test UI, a function call including start and end position, speed, acceleration and deceleration as parameters had to be typed into a command line interface. Here the operator's mental model needs to be incorporated in end-user interfaces. During the shooting he had to adjust the focus manually several times, which was visible in the recorded material. Therefore a motor-driven remote *follow focus*⁵ is a necessary addition. However, professional units are again generally expensive and do rarely offer open access. The do-it-yourself scene provides examples of Arduino-based systems [2]. Such an implementation can be set up for wireless control and would work together with our setup. In response to this finding, we built an Arduino-based remote follow focus. Its parts and sources are also provided in [20].

⁵ A device attached to the camera lens that physically manipulates the focus ring and thus the position of the focal plane. It is often motor-driven and remote-controlled.

Surprisingly, we also found that the people we recorded behaved more *naturally* in front of the camera, which was also confirmed by the operator. Due to the remote control and the automatic bouncing mode, we could remain at a distance from the setup. We had never considered any effects on people in front of the camera before the shooting.

Overall, we see a strong indication that our prototyping platform generally meets the requirement of high quality *smooth camera motions* and of *system stability* for field use we had set out to test. Even though we are aware that it does not reach the level of sophistication of professional equipment, we believe it can still serve as a research platform for new user interfaces prototypes and field observation. During the recording, the slider was indeed controlled wirelessly. However, this was done by a second trained person using a command line interface on a laptop holding it with one hand and typing with the other. We focused on evaluating the collected hardware requirements and not yet the user interface requirements. The user interfaces for end-user control presented below were designed based on insights gained from this first field evaluation.

5 Controlled Experiment on the Effects of a Low Degree of Automation on Workload and Control

In order to meet the requirement of displaying the camera stream, we chose to implement the user interfaces on a tablet capable of this task. We implemented a *touch-based* UI that used the whole screen as an input area for direct control on the camera stream. This design led to less occlusion of the stream by visual interface components and can be extended by further image-based control techniques. We compared this design alternative to a status quo *software joystick* that served as a baseline condition for remote control. Both remote control interfaces were also compared to full *manual control*, a human control baseline without motorization. The human baseline condition allowed us to interpret the data collected beyond a relative comparison between both remote control conditions.

5.1 Measurements

The degree of automation (DOA) in a systems design can be located within a spectrum ranging from full human control to full system control. As presented by Miller and Parasuraman [28], design decisions on this scale are characterized by a trade-off between the reduction/increase of workload and the in-/decrease of the results' predictability (Fig. 3, left). With an increased degree of automation usually comes a decrease in predictability of the results, and with it a decreased sense of control. On this basis we chose a low degree of automation and determined its effects on *workload* and *sense* of control in this experiment. As work by Wen et al. [35] suggests, the quality of the results can itself be affected by the perceived level of control. We hence added a *quality* of control measurement. As no standardized tasks for measuring quality of control with cinematographic interfaces have emerged so far, we adapted a method used in the evaluation of

automotive user interfaces established by Verster and Roth [33]. This method uses the standard deviation of lateral position (SDLP) to determine a driver’s performance. Here, the deviation from the center of the lane is measured continuously while driving. In our study, we used the Rule of Thirds as a basis and continuously determined the deviation from the targeted third (Fig. 3, right).

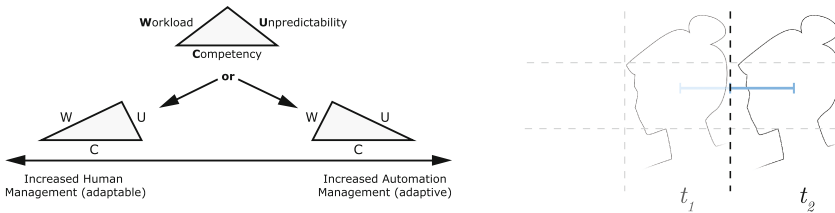


Fig. 3. *Left:* The trade-off between workload and unpredictability as described in [28], *right:* Adaption of the SDLP [33] measurement to the Rule of Thirds principle

5.2 Participants

For the study we recruited 18 participants (14 male). The average age was 24, with ages ranging from 21 to 31. Prior knowledge in tools for camera motion was reported by 4 participants.

5.3 Study Design

Each participant was asked to perform the task of following a person with the camera in movement direction while framing the person at the first third in the direction of the movement (Rule of Thirds technique). The three levels of the independent variable for *interaction technique* were full *manual control* (no motion control, human baseline), *software joystick* (motion control, remote control baseline) and *touch-based control* directly on the camera stream (motion control, touch-based remote control). For *manual control*, the slider carriage needed to be manipulated physically by the participants to move the camera and to frame the person. In the other conditions the slider carriage was driven by a motor and needed to be controlled via a remote control user interface offering continuous control options. In a within-subjects design each participant executed all conditions. The order was counter-balanced based on a Latin Square design.

5.4 Apparatus

An unmotorized slider, for the manual control condition, and a motorized slider, for the other conditions, were mounted on tripods at the same height and placed in front of each other. To provide the participants with the same delay that would

appear in the video stream in the remote control conditions, the video stream was also displayed in this condition. A smartphone was therefore mounted on top of the manual slider carriage to display it. The stream was recorded by a DSLR camera (Canon EOS 60D) mounted on the carriage of the motorized slider. The camera was connected to a video encoder (Teradek Cube 255) via HDMI cable situated on top of the camera. The encoder provided an RTSP video stream of the camera image via WiFi and recorded the stream for post-hoc evaluation. The stream was also presented on the tablet used for the *software joystick* and *touch control* interfaces.

5.5 Procedure

First the participants were welcomed and informed about the study and how their recorded data was handled. They then were handed out a declaration of consent. After declaring consent, a demographic questionnaire was handed out and after its completion, the Rule of Thirds framing task was explained. An example video of the expected results was presented.

The order of the conditions was counterbalanced with a Latin-Square design in order to avoid learning effects. For each condition, the task was executed ten times. For each trial, the video material was recorded for the analysis of *quality of control*. After each condition the participants filled out an extended version of the NASA Task-Load-Index (TLX) [19] questionnaire to determine *workload* and *sense of control*. To determine the latter, the original TLX questionnaire was extended by one item. The question we added was '*How much did you feel in control during the task?*'. The wording of the question was taken from the sense of control scale developed by Dong et al. [15]. Here, the authors propose a 6-point rating scale as the response format. In order to minimize effort and confusion for the participants, we decided to stay consistent with the 20-point scale format used in the TLX. After carrying out all objectives, a semi-structured interview regarding the presented conditions was conducted.

5.6 Results

We conducted Shapiro-Wilk Tests for the data collected on *workload*, *sense of control* and *quality of control*. They showed significance for multiple conditions. In consequence, a normal distribution across all of the data cannot be assumed. We therefore only used non-parametric tests (Friedman's ANOVA and Wilcoxon Signed-Rank) to test for statistical significance. A Bonferroni correction with a value of $\alpha^* = .016$ was applied to account for pairwise comparisons. The post-hoc comparisons were only conducted after a significant main effect was found.

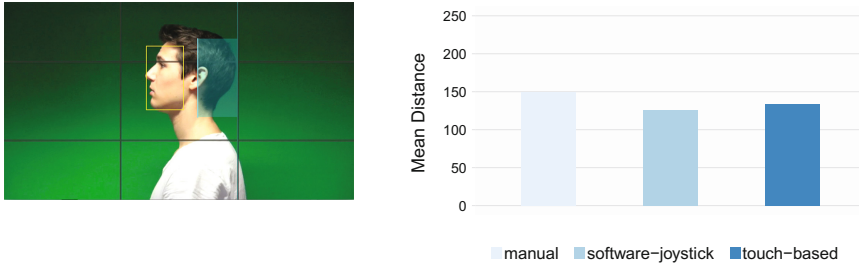


Fig. 4. *Left:* Determining the distance (blue) between the tracking target (yellow) and the first third in movement direction automatically, *right:* Mean distances to the target (Color figure online)

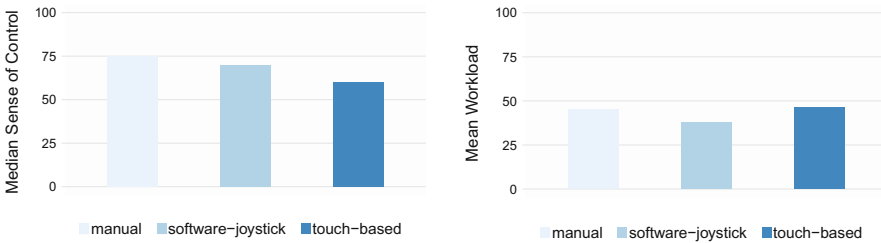


Fig. 5. *Left:* Results for *sense of control*, *right:* Results for the occurring *workload*

Quality of Control. The data determining *quality of control* was extracted from the recorded video material after the study. We developed an analysis tool to detect the face of the person walking by in each frame thus further determining the person’s center (also available electronically [20]). The distance from the first third in movement direction to the person’s center (blue area in Fig. 4, left) in pixels (px) was logged. The resulting data was analyzed with the Friedman test. No significant main effect could be identified ($\chi^2(2) \leq 3.44, p \leq .18$) with mean distance values of 149.94 px (SD ≤ 54.4) difference for *manual control*, 126.19 px (SD ≤ 23.19) for *software-joystick* and 133.03 px (SD ≤ 30.76) for *touch control* (Fig. 4, right).

Sense of Control. For analyzing the self-reported data regarding *sense of control*, we also used the non-parametric Friedman test as the use of parametric tests on self-reported data in rating-scale format is controversial as pointed out by Carifio and Perla [7]. Here also, no significant effect was found ($\chi^2(2) \leq 5.03, p \leq .081$), with median values of 75 for *manual*, 70 for *software-joystick* and 60 for *touch* (Fig. 5, left).

Workload. We determined the *workload* using the TLX for each user interface. Analyzing the overall workload for the different UIs with the Friedman test,

we found no significant main effect ($\chi^2(2) \leq 7.00$, $p \leq .03$). With mean values of 45.09 (SD ≤ 15.18) for *manual* control, 37.78 (SD ≤ 13.47) for *software joystick* and 46.16 (SD ≤ 13.06) for *touch control* (Fig. 5, right).

5.7 Discussion

Based on the data we collected, we could not identify any significant negative effects on the *sense* or the *quality* of control due to the use of remote-controlled tools incorporating a low degree of automation. There was also no significant influence on *workload* when using remote-controlled tools. This seems to be in line with prior findings as presented by Miller et al. [28]. We further examined the data on workload more closely and found that in the dimension of *physical demand* there is a rather large difference in median values of 62.5 for *manual*, 15 for *software-joystick* and 30 for *touch-based*. When inspecting only this dimension, the Friedman test results in a significant difference ($\chi^2(2) \leq 29.6$, $p \leq .001$) between the conditions. The pairwise comparisons revealed significant differences between all conditions (with $p \leq .003$ or less). So, when only concerned with physical demand, we observed that the remote-controlled tools could lead to a decrease in workload. However, the difference in this dimension was not strong enough to influence the participants' overall experienced workload significantly when measured with the TLX questionnaire.

6 Controlled Experiment on the Effects of a Medium Degree of Automation and the Review of Results

In the previous study, no negative effects on *sense* of control became apparent, which was surprising to us. Also, no effects on *quality* of control were observed that could have influenced perceived control indirectly. This could be attributed to multiple causes: the low degree of automation, a too coarse measurement tool or to the unawareness of the consequences on the quality of control by the participants. Regarding the precision of the evaluation tool, a single questionnaire item measured only after one trial might not be sensitive enough to show an effect in the analysis. Additionally, when examining the recorded material, we found that shaking and jerky motion was much more visible in the manual condition, however this had no effect on the sense of control reports. As there was no reviewing of the material, the participants potentially could not estimate the quality of control themselves properly. Providing the possibility to review the results - as usual in cinematographic practice - might influence their perception of control (similar to the findings of Wen et al. [35]) and provide a more externally valid result. To better understand the effects on perceived control, we conducted a second study addressing the issues mentioned above.

6.1 Conditions

Addressing the single degree of automation issue, we now used two designs with a low and medium degree of automation. For the low degree condition, we again

used the software-joystick interface of the prior study. For the medium degree condition we prototyped a UI that used keyframe selection as an input technique. The user interface used a Through-The-Lens [23] approach. The participants could see the scene on the tablet and use it as a viewfinder. They could move the tablet freely and select certain positions as keyframes. The motion-controlled slider would then interpolate a motion path between all chosen keyframes. For the implementation of such a technique, known points of reference are necessary. These can be provided, e.g., via optical tracking or by synchronizing the tablet with a virtual model of the study room. These models consequently need to be updated when the tablet is moved, e.g., through the tracking of markers or the accelerometer data of the tablet. Such an implementation is technically complex as it needs to handle noisy data and requires low latencies. As we were mainly interested in the effects on the perception of our participants, we prototyped it in a wizard-of-oz style. To create the illusion of an automated system capable of this functionality, we asked the participants to select a pre-defined set of keyframes by showing sample images during the explanation of the study task. The motion control tool was accordingly pre-programmed so that the “selected” keyframes were matched by the resulting motion.

6.2 Measurements

To address the issue of a too coarse measurement tool, we now used a visual analog scale to ask for the preferred level of control. The captions on each side of the visual-analog scale read “I control the system and the results manually” and “The system controls the results”. Additionally, we also increased the overall number of data points by taking multiple measurements during the study trials. As we used a wizard-of-oz approach with a pre-defined result, we did not collect data on the quality of control as in the prior study.

6.3 Participants

We recruited 12 participants (8 male) for the study. The average age was 25, with ages ranging from 21 to 32. No prior knowledge in tools for camera motion was reported by the participants.

6.4 Study Design and Procedure

We first welcomed the participants and introduced them to the study procedure. Then we handed out a declaration of consent to the participants. Having declared consent, they were given a detailed explanation of the study conditions. Each participant was then assigned to one of two groups. The groups were exposed to the conditions in counterbalanced orders. Before we started the first trial, we took a measurement on the preferred level of control to collect a baseline. Then the participants were exposed to both conditions. After finishing the trials we took a further measurement on the preferred level of control. Now the participants could determine their preference in reference to the varying degrees of

automation. Then we gave the participants the possibility to review the results and again took a measurement. This time we determined the preferred level of control in reference to their perception of the quality of control in each condition. This addressed the missing review opportunity mentioned earlier. After all trials and measurements the participants were debriefed and thanked for their participation.

6.5 Results

We conducted a Shapiro-Wilk Test for the collected data. It showed no significance and therefore, a normal distribution of the data can be assumed. To stay consistent with our prior analysis, however, we again used Friedman's ANOVA to test for significance. The test indicated no significant differences between the 3 measurements ($\chi^2(2) \leq .359, p \leq .836$) with mean values of 40.28 (SD ≤ 18.26) for the baseline measurement, 36.58 (SD ≤ 25.38) for the measurement after exposing participants to the conditions and 39.33 (SD ≤ 23.21) after reviewing the results (Fig. 6).

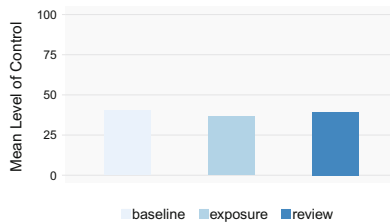


Fig. 6. Results on the measured preferred level of control

7 Reflecting on Our Mixed-Methods Design and Evaluation Approach and Its Results

The camera is generally moved synchronously by multiple operators following a defined workflow with little room for errors. A similar situation – collaborative work in established practices with little room for error – was found by Mackay and Fayard in [24]. Reflecting on their design process, they proposed a framework for HCI research. They recommend a triangulation between *theory*, the design of *artifacts* and *observation*. As HCI is an interdisciplinary field, the application of only certain methods threatens the generalizability or validity of the findings. In lab studies, conditions and variables can be controlled, however the situation is artificial and users might behave unnaturally. In contrast, users behave naturally in field studies, but it is hard to establish cause and effect relationships as influencing factors can interfere. We also used such a mixed-methods approach applying various techniques in our design and evaluation process (Fig. 7). To provide an overview of all of the applied methods, we provide a summary in Table 2.

For the human-automation interplay, user interface design and cinematography, *theory*, design guidelines as well as approaches can already be found in related work [21,28,29]. We therefore mainly focused on the development of a physical prototype, i.e., the design of an *artifact* with following *observations*. Addressing the challenges mentioned in Sect. 1 and the issues mentioned in Sect. 3, we built a motion-controlled camera slider with wireless remote control as a prototyping and research platform. Its use for prototyping is (a) *manageable* and *affordable*. Also (b) the occurring workload in control can be handled even by *non-experts* (non-exclusiveness). Through the wireless interface, it can easily (c) serve as a research platform as various user interfaces (e.g., mobile devices, gestural recognition devices etc.) or computer vision-based systems can be connected and evaluated. Additionally, it can (d) be *transported easily* and therefore is suited for lab as well as field evaluations.

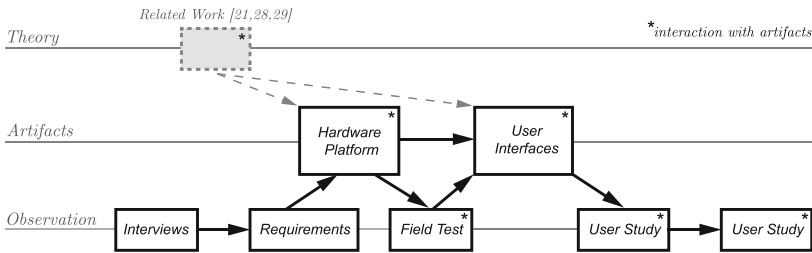


Fig. 7. Our work located within the framework of [24] incorporating control strategies as proposed by [21]. Black (solid frame) items represent the scope of this work.

As our physical implementation is mainly based on open source technology, we also provide it for reproduction and customization. Therefore the changing needs of researchers and practitioners can be met in varying contexts beyond the traditional scope of cinematography, as shown by the examples presented in Sect. 2. Identifying user roles, hardware and feature requirements with experts in a user-centered fashion helped us to minimize the risk of possible user rejection due to poor design or build quality in the implementation of our prototype. Experts in cinematography are often used to high-end equipment and thus bringing a prototype to the set has to be considered with caution. This was reflected by the fact that only after the first session and the screening of the results we were granted further access to the expert users and field environment and consequently were able to gather our insights.

Compared to virtual camera control, physical camera control faces different challenges due to in-situ constraints. Therefore it is hard to apply findings from virtual environments to the physical world. Our system is designed for application in the physical world and can be used in lab environments and in field studies alike. Sometimes, natural interaction processes and phenomena such as *unexpected use* emerge only in-the-wild. This is described by Marshall and colleagues [27]. These phenomena can give further insight into the users' actions,

reasoning and experiences. In our case, one surprising and mentionable field insight was that untrained people tend to behave more naturally when filmed by an autonomously operating or remotely operated system. We also observed that the operators would ask to explore more variations of the same shot at different speeds, to have more options in choosing the fitting material for the rhythm of the cuts and the soundtrack in post-production.

We additionally examined two particular research questions in controlled studies. First we wanted to gain insights on how workload as well as sense and quality of control were affected by the introduction of a low degree of automation compared to a full human baseline. In conclusion, we could not find any negative effects regarding either sense or quality of control in our data. Although we could find a reduction in workload when only concerned with physical effort, this did not significantly decrease the overall workload score in our data set.

Table 2. The methods we used in our mixed-methods design and evaluation approach

Method	Results
Online survey (N = 3)	User roles from practitioners
Expert interview (N = 1)	Hardware and user interface requirements
Expert evaluation (N = 1)	Testing hardware implementation, Contextual inquiry with field insights, Iteration of user interface requirements
Controlled user study 1 (N = 18)	Data on the effects of introducing automation
Controlled user study 2 (N = 12)	Data on the effects of introducing a review phase

Given our experiences from the field evaluations, we were surprised *not* to find a difference. We had expected that a missing screening of the results might have affected the reports, in particular, on sense of control. In the full manual condition more and more noticeable jerky motions were observable in the recorded material. Participants however might have been unaware of it without a review phase. We hence conducted a second controlled user study examining the effects of adding a review phase to the evaluation process. Based on our collected data, we concluded that the introduction of a review phase had no significant effect on the participants' perception of control even when increasing the level of automation. To mimic the recording practice we observed on set and given that jerky motions might only be discovered after a recording, we would still recommend its integration in the evaluation process in field as well as laboratory environments.

Besides the general triangulation approach proposed by Mackay and Fayard, Shneiderman et al. [30] propose the use of a cascade of evaluation techniques in particular when evaluating creativity support tools (CST). We consider camera motion control tools to be a peculiar kind of CSTs and as such, their evaluation should not be conducted on performance measures alone. As pointed out in [30],

performance measures might be part of the evaluation, but field observation and other quantitative measurement dimensions should also be considered as for example proposed by Cherry and Latulipe [11] in their Creativity-Support Index (CSI) questionnaire.

Expanding automation in the domain of creative expression seems paradoxical at first. Delegating camera motion to machines might result in a more robotic aesthetic than in manual operation as found in related work [10]. However, such tools can add to the vocabulary and expressiveness of the discipline as in time-lapse, high-speed or aerial shots, which can hardly be controlled manually. Additionally, smart, easy to use tools offering fast implementation of ideas that might lead to more exploration.

Evaluating and quantifying such aspects is a limitation of the presented work. With sense of control, we did consider an experience measurement beyond the performance measures of workload and quality of control. In our field observations, we also observed how the use of our tools influenced expressiveness and exploration of various speed settings in order to shape different versions for post-production. Expressiveness and exploration are, for example, dimensions measured by the CSI. We did not investigate these or any other dimensions covered by the CSI in detail.

Given the early development stage of our prototype and the controlled study design and goal-oriented study task (adapted SDLP), the environment was not particularly well suited for determining aspects such as “results worth effort” or “enjoyment”. These are clearly important aspects that CSTs should support and that should be evaluated. However, we believe they also require an open-ended task and potentially an even more stable system.

8 Conclusion

For user interface design and prototyping in cinematographic motion control camera systems we identified a number of challenges. For example, we found that translating new ideas into working prototypes can be hard as expertise in multiple fields is required. Bringing together the expertise of camera operators, a mechanical engineer and computer scientists, we contribute a tool that can be used for prototype development. In expert interviews we identified hardware and user interface requirements. We integrated those in our implementation of a motion-controlled camera slider. It serves as a research platform, as through the wireless remote access that it provides, any type of user interface integrating Bluetooth, such as mobile devices, gesture recognition or wearable devices can be connected. In five on-assignment shootings with an expert we found strong indication that our system fulfills the collected requirement and thus is qualified for field use. We further reported two user studies ($N = 18$, $N = 12$) examining the effects of different degrees of automation on sense and quality of control of the participants. We could not determine any negative effects caused by the use and automated motion control tool, even when compared to a full manual control

baseline. We also examined the effect of including the review of results in the evaluation process and found no indication that its integration led to significant differences in self-reports on the preferred level of control. Providing the system as open source, we encourage its reproduction, customization and extension by researchers and practitioners.

References

1. Freely MIMIC for MöVI. <http://freelysystems.com/mimic>
2. Soffer/Follow-Focus · GitHub. <https://github.com/soffer/Follow-Focus>
3. Star Wars Visual Effects, from AT-ATs to Tauntauns. <https://www.youtube.com/watch?v=mIlyk7KQe-s>
4. STYPE KIT - 3D Virtual Studio and Augmented Reality System Stype Grip. <http://www.stypegrip.com/stype-kit>
5. Bowman, D.A., Kruijff, E., LaViola Jr., J.J., Poupyrev, I.: 3D User Interfaces: Theory and Practice. Addison-Wesley, Westford (2004)
6. Byrne, K., Proto, J., Kruysman, B., Bitterman, M.: The power of engineering, the invention of artists. In: McGee, W., de Ponce Leon, M. (eds.) *Robotic Fabrication in Architecture, Art and Design 2014*, pp. 399–405. Springer, Cham (2014). doi:10.1007/978-3-319-04663-1_30
7. Carifio, J., Perla, R.J.: Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *J. Soc. Sci.* **3**(3), 106–116 (2007)
8. Carr, P., Mistry, M., Matthews, I.: Hybrid robotic/virtual pan-tilt-zoom cameras for autonomous event recording. In: *Proceedings of the 21st ACM International Conference on Multimedia, MM 2013*, pp. 193–202. ACM (2013)
9. Chen, J., Carr, P.: Autonomous camera systems: a survey. In: *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, Québec, Canada*, pp. 18–22 (2014)
10. Chen, J., Carr, P.: Mimicking human camera operators. In: *IEEE Winter Conference on Applications of Computer Vision, WACV 2015*, pp. 215–222. IEEE (2015)
11. Cherry, E., Latulipe, C.: Quantifying the creativity support of digital tools through the creativity support index. *ACM Trans. Comput. Hum. Interact. (TOCHI)* **21**(4), 21 (2014)
12. Christie, M., Hosobe, H.: Through-the-lens cinematography. In: Butz, A., Fisher, B., Krüger, A., Olivier, P. (eds.) *SG 2006. LNCS*, vol. 4073, pp. 147–159. Springer, Heidelberg (2006). doi:10.1007/11795018_14
13. Christie, M., Normand, J.M.: A semantic space partitioning approach to virtual camera composition. *Comput. Graph. Forum* **24**(3), 247–256 (2005)
14. Christie, M., Olivier, P., Normand, J.M.: Camera control in computer graphics. *Comput. Graph. Forum* **27**(8), 2197–2218 (2008). Blackwell Publishing Ltd
15. Dong, M.Y., Sandberg, K., Bibby, B.M., Pedersen, M.N., Overgaard, M.: The development of a sense of control scale. *Front. Psychol.* **6**, 1733 (2015)
16. Fitts, P.M.: *Human Engineering for an Effective Air-Navigation and Traffic-Control System*. National Research Council, Division of Anthropology and Psychology, Committee on Aviation Psychology (1951)
17. Gebhardt, C., Hepp, B., Nägeli, T., Stevsic, S., Hilliges, O.: Airways: optimization-based planning of quadrotor trajectories according to high-level user goals. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI 2016*, pp. 2508–2519. ACM, Santa Clara (2016)

18. Gleicher, M., Witkin, A.: Through-the-lens camera control. In: Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1992, pp. 331–340. ACM, New York (1992)
19. Hart, S.G.: NASA-task load index (NASA-TLX); 20 years later. *Proc. Hum. Factors Ergon. Soc. Ann. Meet.* **50**, 904–908 (2006)
20. Hoesl, A.: CameraMotion · GitLab. <https://gitlab.lrz.de/lmu08360/CameraMotion>
21. Hoesl, A., Wagner, J., Butz, A.: Delegation impossible? Towards novel interfaces for camera motion. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA 2015, pp. 1729–1734. ACM, April 2015
22. Hulens, D., Goedemé, T., Rumes, T.: Autonomous lecture recording with a PTZ camera while complying with cinematographic rules. In: Proceedings of the 2014 Canadian Conference on Computer and Robot Vision, CRV 2014, pp. 371–377. IEEE Computer Society, Washington, DC (2014)
23. Lino, C., Christie, M., Ranon, R., Bares, W.: The director’s lens: an intelligent assistant for virtual cinematography. In: Proceedings of the 19th ACM International Conference on Multimedia, MM 2011, pp. 323–332. ACM, Scottsdale (2011)
24. Mackay, W.E., Fayard, A.L.: HCI, natural science and design: a framework for triangulation across disciplines. In: Proceedings of the 2nd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 223–234. ACM (1997)
25. Marchal, D., Moerman, C., Casiez, G., Roussel, N.: Designing intuitive multi-touch 3D navigation techniques. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013. LNCS, vol. 8117, pp. 19–36. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-40483-2_2](https://doi.org/10.1007/978-3-642-40483-2_2)
26. Marchand, E., Courty, N.: Image-based virtual camera motion strategies. In: Fels, S., Poulin, P. (eds.) Proceedings of the Graphics Interface 2000 Conference, GI 2000, pp. 69–76. Morgan Kaufmann Publishers, Montréal (2000)
27. Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., Davies, M.: Rethinking ‘Multi-user’: an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2011, pp. 3033–3042. ACM, New York (2011)
28. Miller, C.A., Parasuraman, R.: Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. *Hum. Factors J. Hum. Factors Ergon. Soc.* **49**(1), 57–75 (2007)
29. Nielsen, J.L.: Camera Movement in Narrative Cinema: Towards a Taxonomy of Functions. Århus Universitet, Institut for Informations- og Medievidenskab (2007)
30. Shneiderman, B., Fischer, G., Czerwinski, M., Myers, B., Resnick, M.: NSF Workshop Report on Creativity Support Tools. Technical report, Washington, DC, USA (2005)
31. Stanciu, R., Oh, P.Y.: Designing visually servoed tracking to augment camera teleoperators. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IRDS 2002, vol. 1, pp. 342–347. IEEE (2002)
32. Tognazzini, B.: The “Starfire” video prototype project: a case history. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1994, pp. 99–105. ACM (1994)
33. Verster, J.C., Roth, T.: Standard operation procedures for conducting the on-the-road driving test, and measurement of the standard deviation of lateral position (SDLP). *Int. J. Gener. Med.* **4**(4), 359–371 (2011)

34. Ware, C., Osborne, S.: Exploration and virtual camera control in virtual three dimensional environments. In: ACM SIGGRAPH Computer Graphics, vol. 24, pp. 175–183. ACM (1990)
35. Wen, W., Yamashita, A., Asama, H.: The sense of agency during continuous action: performance is more important than action-feedback association. PLoS One **10**(4), e0125226 (2015)
36. Zhang, Z., Liu, Z., Zhao, Q.: Semantic saliency driven camera control for personal remote collaboration. In: IEEE 10th Workshop on Multimedia Signal Processing, MMSP 2008, pp. 28–33. IEEE (2008)