TrackLine: Refining touch-to-track Interaction for Camera Motion Control on Mobile Devices

Axel Hoesl(✉), Sarah Aragon Bartsch, and Andreas Butz

LMU Munich, Munich, Germany
{axel.hoesl,sarah.aragon.bartsch,andreas.butz}@ifi.lmu.de

Abstract. Controlling a film camera to follow an actor or object in an aesthetically pleasing way is a highly complex task, which takes professionals years to master. It entails several sub-tasks, namely (1) selecting or identifying and (2) tracking the object of interest, (3) specifying the intended location in the frame (e.g., at 1/3 or 2/3 horizontally) and (4) timing all necessary camera motions such that they appear smooth in the resulting footage. Traditionally, camera operators just controlled the camera directly or remotely and practiced their motions in several repeated takes until the result met their own quality criteria. Automated motion control systems today assist with the timing and tracking subtasks, but leave the other two to the camera operator using input methods such as touch-to-track, which still present challenges in timing and coordination. We designed a refined input method called TrackLine which decouples target and location selection and adds further automation with even improved control. In a first user study controlling a virtual camera, we compared TrackLine to touch-to-track and traditional joystick control and found that the results were objectively both more accurate and more easily achieved, which was also confirmed by the subjective ratings of our participants.

Keywords: Camera motion · Motion control · Image-based control · User interface · User-centered design

1 Introduction

New technologies such as drones, gimbals or industrial robots have substantially advanced physical cinematographic camera motion and motion control within the last decade. These systems provide smooth camera motion and image stabilization. They offer more accuracy and reproducibility to experts and simultaneously lower the entrance barrier for novices. As these systems are motor-driven, they can be operated remotely, often by manipulating input hardware or virtual interface elements on the touch screen of a mobile device.

In fact, mobile devices are particularly appealing as they offer additional functionality, such as reviewing the video stream of live or recorded material and

R. Bernhaupt et al. (Eds.): INTERACT 2017, Part II, LNCS 10514, pp. 283–292, 2017.
DOI: 10.1007/978-3-319-67684-5_17

quick editing of sequences. They also allow to flexibly control different systems with one controller or to share material with others. On the other hand, touch devices also introduce their own challenges, most prominently, a loss in control precision. Particularly the lack of haptic and kinesthetic feedback makes the already complex task of camera control even more challenging.

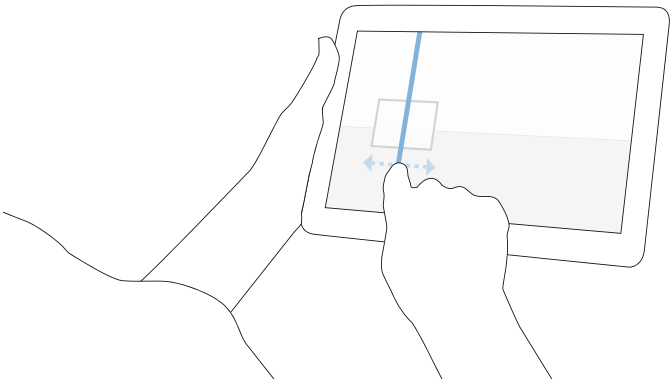


Fig. 1. Users can predefine the axis at which a moving object should be framed by dragging the TrackLine (blue) at the desired position (Color figure online)

Manufacturers have introduced automatic functions, such as object tracking using computer vision (CV) to overcome some of these new challenges, sometimes allowing image-based motion control directly on the video stream. Instead of continuously controlling the movement with a hard- or software joystick, operators can now specify the expected results (e.g. the framing of a moving object in the image). A motor-driven gimbal system can then not only keep the camera steady, but also maintain framing when following the moving object[1].

For such image-based control, systems often use the touch-to-track (TTT) approach for the selection of an object to be tracked and followed. With TTT, users tap on the object or person in the video stream and the system then continuously adjusts the camera position to keep the selected object at the same position within the frame. This design entangles object selection, framing and timing in one interaction, making it fast, but also prone to errors as the touch interaction needs to be timed precisely. The object might be moving out of the frame without being selected or the user might not be able to tap on the object at the right moment. In this case, the user needs to perform a select-and-correct move to adjust the framing position: with a drag-and-drop gesture on the touch screen, the object is moved to its correct framing position. As the correction is performed while recording, the resulting film material can only be used after the select-and-correct move has been carried out. Performance with this manual selection of a moving object can also be expected to decrease for faster moving targets [6] as in car commercials, sports-broadcasting or high-speed recordings.

[1] Example system: http://www.vertical.ai/studio.

1.1 Contributions

To overcome the issues of existing image-based motion control methods, we developed TrackLine as an alternative method, untangling the interactions of a touch-to-track design. TrackLine (Fig. 1) lets operators define the desired tracking position in advance and delegates the correct timing to an assisting system. A vertical line, displayed on top of the video stream, serves as a motion trigger. It can easily be positioned by drag and drop gestures before the recording is started. As soon as a moving object intersects with the line in the image space, the camera starts moving automatically, framing the object at the predefined position. By defining the desired tracking position in advance, select-and-correct moves can be avoided and the selection of fast objects is no longer tied to human reaction time. We compared our approach to existing approaches, namely software joystick and touch-to-track. Our results show that, TrackLine is more *efficient* (fewer retakes) and more *precise* (smaller distance from intended position). In addition, it was perceived as easy and effective to use as well as quick to learn by our participants.

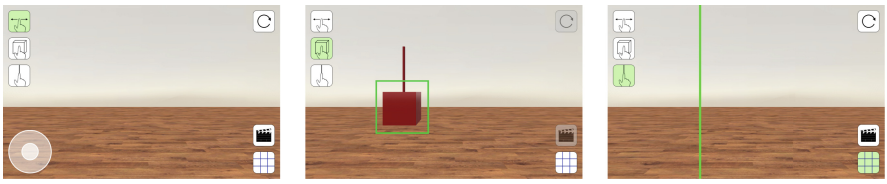
2 Related Work

Focusing on physical camera motion, Chen and Carr [3] presented an in-depth survey on autonomous camera systems. They identified core tasks and summarized twenty years of research-driven tool development and evaluation. Besides traditional cinematography, further work was conducted in automated lecture recording [12,18–20], tele-conferencing [10,21] or event broadcasting [1,2,4,9]. Regarding high-level control in particular, as in our case, the most prominent approaches are Through-the-Lens [7], image-based [14] or constraint-based [8,13] control. Our design uses image-based control on a *virtual* camera, but today's motorized motion control systems already show that results for virtual camera control can be applied to remotely controlled real cameras. To evaluate designs the use of standardized tasks is common. For comparing different techniques in cinematographic tasks however, we found no well-established methodology in the literature. Of course, systems have been evaluated before in their unique ways. In [16] a computer-vision supported crane was evaluated by following a target that was moved by an industrial robot. In [12] a motorized slider was used to move an action figure, which served as a tracking target for a computer vision based panning system. Without standardized tasks at hand, a system's capabilities are also often documented by picture series as in [15] or by referencing a video as in [11]. These approaches, however, are often limited to subjective interpretation. To enable *objective* comparisons with a task that is also native to cinematography, we chose a framing task and measure similar to [12] and [17].

3 Study Comparing TrackLine to the State of the Art

In a user study, we compared our TrackLine approach to the current state of the art, which is the software joystick for continuous and touch-to-track for assisted control, representing the baselines for two different control paradigms (Fig. 2).

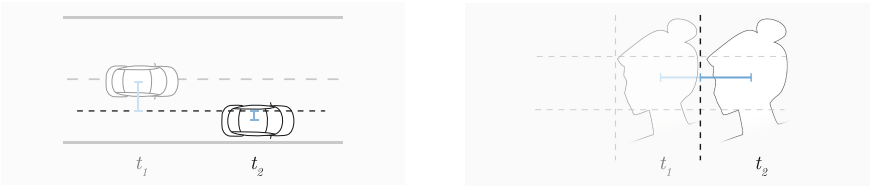
Using the software joystick, the camera could be translated manually along the x-axis, while with TTT, the motion was triggered by tapping on the target at the desired tracking position. With our TrackLine approach, the users had to drag the motion axis at the desired position before starting the recording. For the assisted methods, visual feedback was given by indicating the automated tracking with a green box around the target.



(a) software-joystick (b) touch-to-track (c) TrackLine

Fig. 2. The status quo designs software-joystick and touch-to-track (Figs. 2(a) and (b)) and our TrackLine alternative (Fig. 2(c)) as implemented and evaluated (Color figure online)

To test our approach early in the design process, we used a virtual camera and environment for a prototypical implementation. The concept was implemented in Unity 5.3. running on an off-the-shelf Android tablet.



(a) Deviations (blue) from lane center (b) Deviations (blue) from desired framing

Fig. 3. The SDLP [17] for estimating the quality of control in a driving task (Fig. 3(a)), our adaptation (Fig. 3(b)) to determine *precision* in a cinematographic task (Color figure online)

For the study we wanted to collect quantitative data with a task that is native to cinematography. We therefore adapted an appraoch from the automotive domain, namely the standard deviation of lateral position (SDLP) [17]. The SDLP measure follows the idea that when driving – usually in a simulator – during the course of a study participants will deviate from the center of the lane they are driving on (Fig. 3(a)). These differences from the ideal pathway are recorded and analyzed. The closer the participants' trajectory matches the ideal pathway, the better is their driving performance.

Our adaption is based on the Rule of Thirds[2]. Beyond its aim to provide a visually pleasing framing for the audience, it can also be understood as a

[2] Guideline for image composition where the image space is divided into thirds (horizontally and vertically). Objects of interest are best placed at one of the intersections.

goal-oriented task for the operator. The goal-orientation allows to measure how accurately the goal is achieved. Similar to the SDLP, we use the distance between the moving object's actual and ideal position. The ideal position in our case is the first third in movement direction within the image space (Fig. 3(b)). The smaller the distance from the ideal position, the better we rate the control performance.

3.1 Study Design

We used a within-subjects study design with the independent variables user interface (3 levels) and task (3 levels). To avoid learning effects both variables were counter-balanced with a Latin-Square Design.

3.2 Participants

We recruited 12 participants (3 female). The average age was 24, with ages ranging from 21 to 27. Vision was normal or corrected to normal for all. Also, all were familiar with touch screens and 3 acquainted with camera operation.

3.3 Study Tasks

In our study, a horizontally moving object (red cube) should be followed and framed according to the Rule of Thirds. In detail, when the recording was started by tapping on a button, a text countdown of three seconds was displayed on screen, before the cube moved into the scene from the left. The users were asked to frame the cube at 1/3 of the screen, e.g., the first third in movement direction. To find the target position more easily, a thirds grid could be displayed by the participants. Additionally, the center of the cube was marked with an antenna. The camera motion should be stopped when a red signal was presented on the display. To vary the workload, we developed three variations of this task: *direction change* (movement direction changes), *fast object* (object moves at high velocity) and *track&pan* (transition from tracking to panning). Within the latter, the users had to smoothly transition the camera motion from a translation along the x-axis to a rotation around the y-axis when the red signal was shown. For manual control, we therefore implemented a second software joystick. When using the assisted techniques, the users only hat to tap on the background of the screen to trigger the transition.

3.4 Procedure

After welcoming the participants and explaining the structure and the context of the study, a consent form was handed out. Given their consent, the participants were asked to fill out a demographic questionnaire. They were then given an introduction to the application and could familiarize with the controls during a five minute training task in which the cube moved from left to right at moderate speed. Following the training task, the specifics of each task variation were

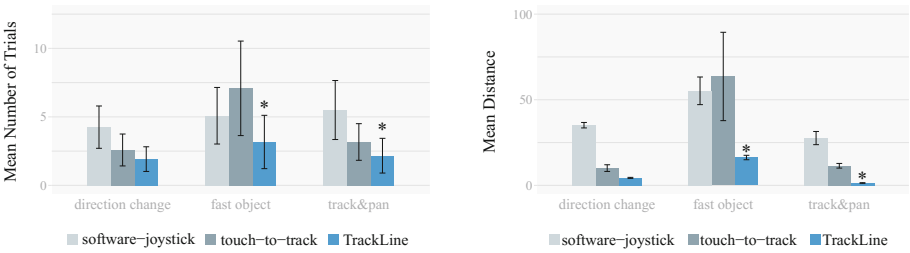
explained in detail ahead of each condition. After finishing each task, the participants were asked to fill out a questionnaire and the next task was prepared. To be consistent to the mobile usage context, we asked the participants to carry out the tasks while standing. The study ended with a final questionnaire and semi-structured interview, asking the users to give a rating of the techniques in direct comparison.

3.5 Measurements

Participants were asked to repeat the task until they did not expect a further increase in performance anymore. They were assessing their results based on their own subjective impression. As a measure for *efficiency*, we thus counted the number of trials. We consider this a relevant measure, as in real world productions a reduced number of trials effectively saves time and money. For estimating *precision*, we used our adaptation of the SDLP described above. We continuously logged the distance of the moving cube to the ideal position.

4 Data Analysis and Results of the User Study

For the data collected on *efficiency* and *precision* the conducted Shapiro-Wilk Tests showed significance. Thus normality of the data cannot be assumed. In the following data analysis we therefore used non-parametric tests (Friedman's ANOVA, Wilcoxon Signed-Rank) to test for statistical significance. Bonferroni correction was applied in post-hoc tests to make up for pairwise comparisons ($\alpha^* = .016$). These were conducted only after a significant main effect was found.



(a) Mean number of trials (b) Mean distances from third

Fig. 4. Results on *efficiency* and *precision* of the studied interfaces

4.1 Efficiency

In number of trials, the *fast object* and *track&pan* tasks showed noteworthy effects. For *fast object*, TTT needed the most trials and for *track&pan* the software-joystick did. For each task one approach would perform better and

therefore seems better suited. However, TrackLine outperformed both interfaces in both tasks. For *fast objects* ($\chi^2(2) = 19.96$, $p \leq .001$), participants did 5.08 trials on average (SE $\leq .06$) with software-joystick, while they did 7.08 (SE ≤ 1.0) with TTT and only 3.17 with TrackLine (SE $\leq .56$). Post-hoc pairwise comparisons showed that using TrackLine resulted in significantly fewer trials than software-joystick ($Z \leq 3.11$, $p \leq .002$, $\eta^2 \leq .81$) and TTT ($Z \leq 2.96$, $p \leq .003$, $\eta^2 \leq .73$). For *track&pan* ($\chi^2(2) = 21.04$, $p \leq .001$), the participants did 5.50 trials on average with software-joystick (SE $\leq .62$), with TTT 3.17 (SE $\leq .39$) and only 2.17 (SE $\leq .37$) with TrackLine. Post-hoc tests indicate that TrackLine needs less trials than software-joystick ($Z \leq 3.08$, $p \leq .002$, $\eta^2 \leq .79$) and TTT ($Z \leq 2.81$, $p \leq .005$, $\eta^2 \leq .66$).

4.2 Precision

For *fast object* ($\chi^2(2) \leq 13.17$, $p \leq .001$), the software-joystick resulted in a 55.22 px distance on average (SE ≤ 8.07), TTT in 63.60 px (SE ≤ 25.80), and TrackLine only in 16.26 px (SE ≤ 1.30). In pairwise comparison, we found participants to be closer the ideal position with TrackLine than with software-joystick ($Z \leq 2{,}98$, $p \leq 0.06$, $\eta^2 \leq .74$) and TTT ($Z \leq 2{,}75$, $p \leq 0.03$, $\eta^2 \leq .63$). For the *track&pan* task ($\chi^2(2) \leq 22.17$, $p \leq .001$), the average distances were 27.62 px with software-joystick (SE ≤ 3.85), 11.45 px with TTT (SE ≤ 25.80) and only 1.34 px with TrackLine (SE ≤ 0.17). Pairwise comparisons reveal that TrackLine led to a smaller distance to the ideal position than software-joystick and TTT (for both: $Z \leq 3{,}06$, $p \leq 0.02$, $\eta^2 \leq .78$). Additionally, it helped avoiding misses that occurred especially with TTT resulting in a larger variance (see error bars in Fig. 4(b)).

4.3 User Feedback

We collected self-reported data on the efficiency, ease of use and comfort via 5-item rating scales. The software-joystick was rated low in efficiency (Mdn = 2) and ease of use (Mdn = 2), but as comfortable to hold (Mdn = 4.5). TTT was rated to be efficient (Mdn = 4) and easy to use (Mdn = 4), but lowest in comfort (Mdn = 3.5). TrackLine was perceived as very efficient (Mdn = 5), easy to use (Mdn = 5) and comfortable to hold (Mdn = 5). In the debriefing interviews, participants pointed out that extension in future work should include fostering exploration and expressiveness of the technique. '*I liked the joystick best, because I had the most control, even if the technique is potentially more imprecise than the TrackLine.*' (P09). But the same participant also acknowledged the resulting jerkiness of the motion '*If this would have been a real recording, it would have become pretty jerky.*' (P09). Participants also felt that the software-joystick occupied more cognitive resources. Three participants stated that it was harder to react to the presented stop signal when using the software-joystick, because they were occupied with focusing on the cube.

5 TrackLine: Effective and Precise – Yet Limited

In our study we observed an expected decrease in *efficiency* and *precision*, especially for fast moving objects. Regarding state of the art designs, depending on the task, one or the other seems preferable. Despite these particularities provoked by the different tasks, our approach could address the performance issues in all cases. It resulted in more precise camera motion when controlled with a touch screen device and helped to avoid select-and-correct moves. Still, the observed performance issues could not be totally avoided, but their frequency and consequences could at least be minimized. Additionally, we found that in the participants' perception the gain in precision comes at a trade-off in expressiveness. The expressiveness of tools for the support of creative tasks such as cinematography is an important aspect. It is thus often tested in the evaluation of such tools, for example with the Creativity Support Index by Cherry and Latulipe [5]. Therefore, the missing expressiveness needs to be addressed in future work. This means in particular, (a) to extend the functionality of the TrackLine approach and (b) to combine manual continuous control elements fostering exploration and expressiveness with a content-based approach providing efficiency and precision.

To assess data and insights early in the design process, we implemented our system in a virtual environment. This environment, of course, provides perfect information, which would be more noisy in real world systems. This is likely to uniformly affect all performance aspects we measured, especially regarding *precision*. The collected data should thus be considered with caution on an absolute level. Our results are mainly suited for a comparison of the design alternatives on a conceptual level. It thus can help to inform the design of real-world implementations. While future work is surely necessary to asses the feasibility of our approach in the wild, we think that given today's existing technology, an enhanced version of our approach could already be implemented on top of established systems. However, such an implementation still needs to address the issues discussed above.

6 Conclusion

To overcome the issues of existing image-based motion control methods, we developed a refinement for high level camera motion control on mobile devices (TrackLine). TrackLine lets operators define the desired tracking position in advance and delegates the correct timing to an assisting system. Our approach counteracts select-and-correct moves and separates performance, especially in selecting fast objects, from human reaction time. To evaluate our design, we compared it to two established techniques (software-joystick and touch-to-track). Our results indicate that with TrackLine, operators are more *efficient* while simultaneously being more *precise* compared to established techniques. In addition to the increased performance, users perceived the technique as efficient, easy to use, quick to learn and comfortable to operate.

References

1. Carr, P., Mistry, M., Matthews, I.: Hybrid robotic/virtual pan-tilt-zom cameras for autonomous event recording. In: Proceedings of the 21st ACM International Conference on Multimedia, MM 2013, pp. 193–202. ACM (2013)
2. Chen, C., Wang, O., Heinzle, S., Carr, P., Smolic, A., Gross, M.: Computational sports broadcasting: automated director assistance for live sports. In: 2013 IEEE International Conference on Multimedia and Expo (ICME), ICME 2013, pp. 1–6. IEEE (2013)
3. Chen, J., Carr, P.: Autonomous camera systems: a survey. In: Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence, pp. 18–22. Québec City, Québec, Canada (2014)
4. Chen, J., Carr, P.: Mimicking human camera operators. In: IEEE Winter Conference on Applications of Computer Vision, pp. 215–222. WACV 2015. IEEE (2015)
5. 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)
6. Chiu, T.T., Young, K.Y., Hsu, S.H., Lin, C.L., Lin, C.T., Yang, B.S., Huang, Z.R.: A study of Fitts' Law on goal-directed aiming task with moving targets. Percept. Mot. Skills **113**(1), 339–352 (2011)
7. 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
8. Christie, M., Normand, J.M.: A semantic space partitioning approach to virtual camera composition. Comput. Graph. Forum **24**(3), 247–256 (2005)
9. Foote, E., Carr, P., Lucey, P., Sheikh, Y., Matthews, I.: One-man-band: a touch screen interface for producing live multi-camera sports broadcasts. In: Proceedings of the 21st ACM International Conference on Multimedia, MM 2013, pp. 163–172. ACM, Barcelona, Spain (2013)
10. Gaddam, V.R., Langseth, R., Stensland, H., Griwodz, C., Halvorsen, P., Landsverk, Ø.: Automatic real-time zooming and panning on salient objects from a panoramic video. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 725–726. ACM (2014)
11. Galvane, Q., Fleureau, J., Tariolle, F.L., Guillotel, P.: Automated cinematography with unmanned aerial Vehicles. In: Christie, M., Galvane, Q., Jhala, A., Ronfard, R. (eds.) Eurographics Workshop on Intelligent Cinematography and Editing. WICED 2016. The Eurographics Association (2016)
12. Hulens, D., Goedem, 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, pp. 371–377. CRV 2014. IEEE Computer Society, Washington, DC, USA (2014)
13. 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, Arizona, USA (2011)
14. 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, Montral, Qubec, Canada (2000)
15. 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)

16. Stanciu, R., Oh, P.Y.: Feedforward-output tracking regulation control for human-in-the-loop camera systems. In: Proceedings of the 2005, American Control Conference, ACC 2005, vol. 5, pp. 3676–3681. IEEE (2005)
17. 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. Gen. Med. **4**(4), 359–371 (2011)
18. Wulff, B., Fecke, A.: LectureSight - an open source system for automatic camera control in lecture recordings. In: 2012 IEEE International Symposium on Multimedia (ISM), ISM 2012, pp. 461–466. IEEE (2012)
19. Wulff, B., Rolf, R.: Opentrack-automated camera control for lecture recordings. In: 2011 IEEE International Symposium on Multimedia (ISM), ISM 2011, pp. 549–552. IEEE (2011)
20. Zhang, C., Rui, Y., Crawford, J., He, L.W.: An automated end-to-end lecture capture and broadcasting system. ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) **4**(1), 6 (2008)
21. 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)