NudgeCam: Toward targeted, higher quality media capture

Scott Carter, John Adcock, John Doherty
FX Palo Alto Laboratory, Inc.
3400 Hillview Ave., Bldg. 4
Palo Alto, CA 94304 USA
(carter,adcock,doherty)@fxpal.com

Stacy Branham
Center for Human-Computer Interaction
Virginia Tech
Blacksburg, VA 24061
sbranham@vt.edu

ABSTRACT

NudgeCam is a mobile application that can help users capture more relevant, higher quality media. To guide users to capture media more relevant to a particular project, third-party template creators can show users media that demonstrates relevant content and can tell users what content should be present in each captured media using tags and other meta-data such as location and camera orientation. To encourage higher quality media capture, NudgeCam provides real time feedback based on standard media capture heuristics, including face positioning, pan speed, audio quality, and many others. We describe an implementation of NudgeCam on the Android platform as well as field deployments of the application.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**]: User Interfaces

General Terms

Design, Human Factors

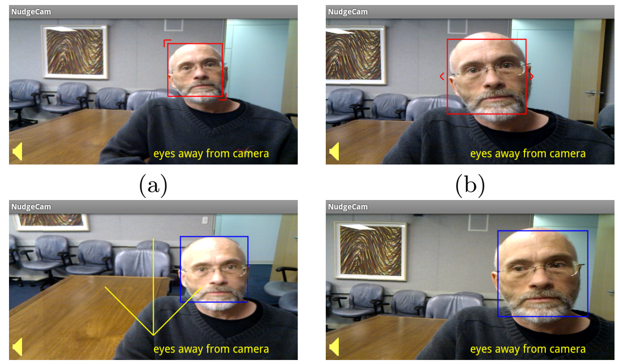
Keywords

Mobile, multimedia, capture

1. INTRODUCTION

The integration of high quality cameras into devices such as digital cameras and cell phones, coupled with the rapid growth of media sharing web sites, has lowered the threshold for media capture and led to rapid growth in photo taking and video recording. There are two methods of approaching the "media overload" problem that has resulted: *post hoc* filtering and *in situ* feedback. Most research [4, 16, 5] as well as commercial [13, 11] efforts fall into the former camp and concentrate on organizing and searching content.

The focus of our work, on the other hand, is in situ feedback. In particular, we are interested in providing guidelines at the point of capture to help users capture better media. Media capture can be improved in two ways. First, it takes a certain amount of skill to record media (especially video) that is palatable – ultimately, no amount of editing can save video that is poorly shot. To help users record higher quality media we can embed well known media capture heuristics into the capture system. Second, in many cases media is captured as part of a specific task or with a specific audience in mind. For example, university public relations workers may want to capture a particular set of photos and videos that can be used to populate a tour guide of their campus. To help users record more targeted media we can support the creation of basic templates, or guides.



(a) (b) (c) (d)

Figure 1: NudgeCam integrates well known capture heuristics, including interview guidelines (e.g., [9]). When a face is detected that is too small, (a), the face is outlined in red with markers at the corners suggesting that the face be increased in the view. When the face is too centered, (b), it is outlined in red with markers indicating that the face should be moved to the side. When the camera is not in the same plane as the subject, (c), a yellow arrow indicates that the user should move the camera to align with the subject. When the face is detected cleanly, (c) and (d), it is outlined in blue. In all cases, a text reminder suggests that the subject not look directly into the camera.

Ultimately, of course, both post hoc and in situ methods can work together to address media overload. However, we believe that in most situations media capture can be improved with appropriate feedback.

To experiment with point of capture feedback, we built NudgeCam, a media capture application for the Android system. The application is designed primarily for two potentially overlapping audiences: third-party template creators and end users. Template creators can configure projects in order to suggest the types of content that end users should capture. NudgeCam also provides suggestions based on well-known media capture heuristics. Importantly, both types of suggestions are just that – suggestions. End users ultimately control the capture process.

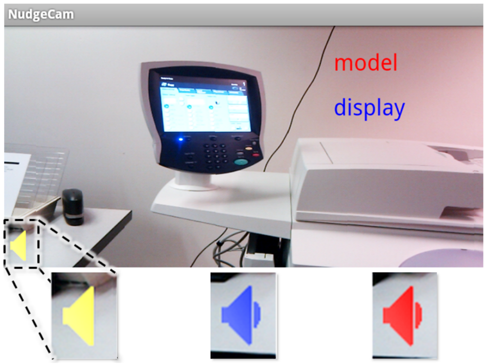


Figure 2: NudgeCam provides real time audio feedback: no audio detected (bottom left), audio detected (bottom center), or audio detected but poor quality (bottom right). While making a recording, tags that the template creator has attached to this demonstration are visible (top right). During the recording the user can tap tags to indicate content that has been included in the capture.

2. IMPROVING MEDIA QUALITY, RELEVANCE, AND USEFULNESS

In their work mapping the design space of systems that direct human action, Heer et al. describe an interactive loop that includes direction, capture, analysis, and feedback [6]. They categorize direction into three categories: show (present a demonstration), tell (relay instructions), and make (provide feedback). We utilize these distinctions to characterize guided capture in the NudgeCam system. Third-party template creators can provide instructions and demonstrations to guide end users to capture particular types of content. Templates consist of a prioritized list of demonstration captures and their associated tags that together act as a to-do list for the end user. The system also provides several different types of largely real time feedback to improve the quality of captured content.

2.1 Show

Template creators use the demonstration media to show end users an example of the type of media to capture. The demonstration need not necessarily be the same media as the recorded media – a photo could be used to show the best spot to take a video and vice versa. Template creators can also modify the order of captured shots, setting priorities on different types of content for end users.

2.2 Tell

NudgeCam provides a number of ways for template creators to instruct end users regarding the content of captured content. First, template creators can add tags to captured content. Tags will appear to end users as reminders listed on the right-hand side of the recording screen. End users can click on tags to indicate that they have captured a particular piece of content. End users are also free to ignore tags altogether. After they are finished recording, NudgeCam associates the captured content with tags that the end user indicated. These tags appear in the list view and are automatically appended to the title of the media file.

The list interface reveals at a glance which demonstrations have media associated with them and which do not – demonstrations with examples are highlighted in blue while those without examples are highlighted in red.

Template creators can also instruct end users via the name of the demonstration clip, or by using the demonstration clip itself to *tell* the user what to record rather than *showing* them an example. Finally, template creators can reorder demonstration content in the list view.

2.3 Make

After capturing a demonstration video, template creators can specify whether to use the device orientation, location, or time for the recording. These have the effect of guiding the end user to match the recording context of the demonstration. To match the camera orientation, NudgeCam presents onscreen arrows directing the end user to match the pan and tilt of the demonstration video. To match location, NudgeCam provides a map view of the demonstration recordings. In the map view, as end users record examples of videos, markers of the demonstrations on the map turn blue to match their counterparts in the list view.

NudgeCam also includes a number of approaches to provide real time and post hoc feedback to end users.

Face Interviewing is a common task when taking videos. According to Liss, when shooting an interview one should "frame the shot with the head and upper chest showing and with the subject off-center to one side or the other." Also, the subject should never "look at the camera," instead focusing on the videographer or elsewhere [9]. To implement this heuristic in NudgeCam, the system continuously detects faces in video frames. The system then calculates the relative position and size of the most salient face. If the face is too centered or too small, the interface overlays red icons indicating the user should shift the camera (Figure 1). NudgeCam also shows a text reminder to the user to remind subjects not to look directly at the camera. Users can click on the text to dismiss it.

Audio In many cases, the audio recorded during a video is as, or more, important than the visual content (e.g., interviews). Currently, NudgeCam analyzes microphone audio and reports three different states: 1) no audio yet detected; 2) audio detected, but too loud or erratic; 3) audio OK (Figure 2). The first category, no audio, can be a useful reminder for end users. For example, many elicitation studies require the user not only to record an event but also to provide their own comments describing why the event was important to them at the time.

Tilt NudgeCam also computes a continuous average of the device's roll (level with respect to the horizon) and provides real time feedback indicating when the device is not aligned.

Erratic motion While making a recording, NudgeCam monitors the device's internal orientation sensors. It computes a running average of the pan and tilt sensors and automatically provides tactile (vibration) and visual feedback if it detects that the user is moving the device erratically. While some amount of camera jitter can be corrected post hoc, if the device is grossly erratic there is simply no content to recover. As Soo notes, this is "one of the most common shooting mistakes people make" [14].

Brightness NudgeCam analyzes the luminosity of each preview frame and displays an icon when the scene is too dark to perceive the target content.

Tags hit End users can tap on tags listed on the recording screen to indicate that content has been included in the capture. Hit tags are automatically associated with recorded content (Figure 2 right). The total tags hit for a set of examples may not cover all of the tags for a demonstration. When this is the case, the demonstration highlight color is darker.

3. PROCESSING CONTENT

In order to provide content based feedback, NudgeCam analyzes multiple streams of data (Figure 3). The main application thread registers with the operating system to receive three types of events, including (1) frames from the camera preview screen, (2) audio amplitude, and (3) location and orientation data. The Android system exposes preview frames in YCbCr 420 SP format. This format is particularly useful for many image analysis tasks since luminance (Y) values can be extracted quickly. The main thread downsamples and saves the frame's luminance channel and forwards the data to image analysis worker threads including face detection, barcode recognition, and mean brightness calculation. The main thread also converts the raw frames to RGB images, one of which is used later as a thumbnail representing the media capture.

Audio amplitude, orientation, and location data can be reported frequently (roughly up to every 10ms). NudgeCam calculates both a running average and running variance of each sensor. When the running average deviates significantly beyond the current state, the state is updated and the view is refreshed. Variance data can be useful to determine sensor reliability. For audio data, variance is also used to detect unusual audio data (such as pops and scratches).

4. APPLICATIONS

We are working with other researchers to develop templates and extend the feature set of NudgeCam to support tasks, including tour guides, diagnostics, interviews, and qualitative field studies. Here we describe two such projects.

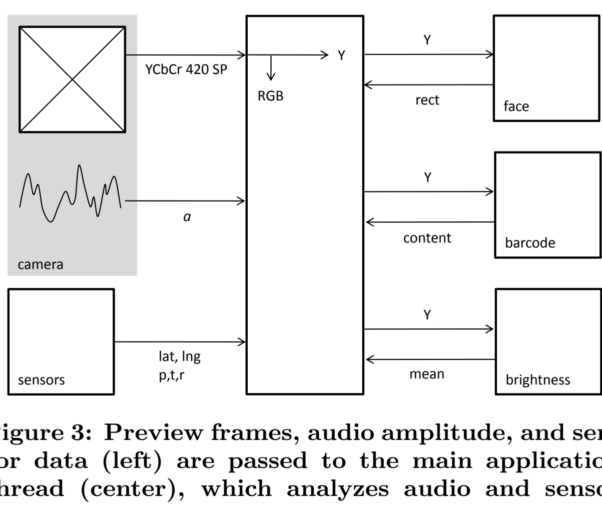


Figure 3: Preview frames, audio amplitude, and sensor data (left) are passed to the main application thread (center), which analyzes audio and sensor data and passes luminance data from the preview frames to worker threads (right).

4.1 Persuasion

Mobile information and communication technologies (ICTs) are the focus of many research projects involving rural communities since they tend to have access to a cellular infrastructure rather than wired networks. While many researchers have deployed ICTs with the goal of helping at-risk and poor communities, most fail because, as Ramachandran et al. argue [12], the technology itself is less important than its use "for persuasion and motivation in order to facilitate change." To explore this issue, Ramachandran et al. deployed a mobile video application to health workers in a poor, rural community. The goal of the project was to have the health workers show videos to community members that encouraged them to visit health clinics. Health workers could also use the device to view testimonial videos for themselves. They could also use the device to capture videos they could replay for potential patients. The researchers found the videos to be an effective method to engage the community and promote better health practices. In particular, they found that the video recordings made by health workers of social influencers (well-respected members of the community) were particularly promising.

We are working with some of the researchers from this study to develop templates for NudgeCam that could be deployed to health workers in the field. These templates can both help the health workers record video more likely to persuade community members and motivate the workers. One way that we are extending NudgeCam to better motivate end users is to integrate a point system, wherein users get points for each demonstration with associated media captures. Points are also awarded for each tag hit per demonstration. Another challenge with this work is that some of the health workers may not be able to read. To address this issue, we are working on a set of icons that could optionally replace or augment text tags.

4.2 Elicitation in diary studies

The diary study is a method of understanding participant behavior and intent in situ that minimizes the effects of observers on participants [1]. Diary studies differ from other field study methods in that researchers are remote from participants and participants control the timing and means of capture. Diary studies can be broken down into those that use media captured by participants as prompts for discussion in interviews (elicitation studies) and those that require participants to answer predefined questions about events (feedback studies). The two methods represent a tradeoff made between accurate recall but burdensome logging (feedback) versus potentially inaccurate recall but unobtrusive logging (elicitation). NudgeCam can help improve recall in elicitation studies by guiding users to capture useful information.

We ran an elicitation study using NudgeCam on two dedicated Nexus One phones in an exploration of design rationale in everyday situations by usability experts. In total, the researcher created three template videos and collected 15 example videos. Three interviews were held with each participant before, during, and after the study. Each interview included questions pertaining to both the goal of the study (design rationales) as well as NudgeCam itself, including perceived usability problems and feature requests. The researcher also documented issues that arose during interviews as participants interacted with NudgeCam. The researcher, who was not involved in the development of NudgeCam, also noted her own experience with the interface.

Overall, the researcher found the NudgeCam valuable as an elicitation device, mainly because it gave her the ability to create a set of template videos and tag them as a guide for study participants. One participant said that he referred back to these template videos regularly, just before recording new examples. Another participant said that the tags made him "feel like [he] had to talk about all of them." He said this caused him to plan out the content of his video a little more thoroughly to make sure he addressed all of them.

Participants identified other potential uses of the NudgeCam, including classroom settings, personal reflection, and documenting important events such as weddings.

5. RELATED WORK

Heer et al. developed guidelines for the design of systems that direct human action [6]. This work provides a useful framework for understanding guided capture as mentioned earlier in this paper. However the work, which extends the Active Capture project [3], focuses on "captur[ing] reusable, annotated media content in a completely automated fashion." NudgeCam, on the other hand, follows an approach more similar to Thaler's and Sunstein's "libertarian paternalism" in which the user is guided but always ultimately controls captured content [15].

Guidance coupled with end user control is also a theme in the MediaTE work [2]. This system uses natural language techniques to suggest a series of video shots based on users' description of the current context (such as where the video is being shot, or who is in the shot) in combination with a desired genre. The suggested shots are intended to satisfy the requirements of a narrative template suitable for automatically generating the final product from the captured shots. Unlike NudgeCam, MediaTE does not base feedback on real time video analysis and allows only coarse-grained feedback on video content; there is no affordance for the user to indicate that they partially fulfilled a suggested shot. Also it is implemented and deployed on a tablet pc or PDA connected to a camera rather than a standalone mobile device.

Kumano et al. developed a similar system using a pc tethered to a camera and a heads-up display that analyzes video frame content in real time to provide feedback regarding erratic pans and zooms [8]. However, the system does not analyze audio data and does not support targeted capture.

Finally, several researchers are working to build computational cameras, which "use unconventional optics and software to produce new forms of visual information" [10, 7]. While both NudgeCam and the computational camera work have similar goals – improving the quality of captured media – they take distinct approaches. The focus of computational camera research is automatic image enhancement, rather than to suggest the type of content that should be in the image. For example, a computational camera could improve the quality of an interview video by automatically adjusting field-of-view, creating HDR (high dynamic range) keyframes, or using a new type of CCD to improve the dynamic range of the video itself. On the other hand, NudgeCam focuses on how the subject appears in the frame. Put another way, the image cannot be improved if the camera is pointing in the wrong direction in the first place. In fact, NudgeCam could potentially work alongside computational camera approaches to improve not only the content but also the clarity and usefulness of captured media.

6. CONCLUSIONS AND FUTURE WORK

NudgeCam is the first mobile tool that supports guided capture by analyzing video and audio content to provide feedback that can improve the quality of captured media while also allowing users to target specific types of content. Even though we consider it still a prototype, we have already found it useful for a wide variety of different scenarios, and we believe the general idea of user-guided capture could have many more applications, including synchronous activities such as video conferencing.

7. ACKNOWLEDGMENTS

We thank Jeremy Pickens and Tony Dunnigan as well as our study participants.

8. REFERENCES

[1] S. Carter and J. Mankoff. When participants do the capturing: The role of media in diary studies. In *CHI '05*, pages 899–908, 2005.

[2] B. Adams and S. Venkatesh. Situated event bootstrapping and capture guidance for automated home movie authoring. In *MULTIMEDIA '05*, pages 754–763, 2005.

[3] M. Davis. Active capture: Automatic direction for automatic movies. In *MULTIMEDIA '03*, pages 602–603, 2003.

[4] A. Girgensohn, J. Adcock, M. L. Cooper, J. Foote, and L. Wilcox. Simplifying the management of large photo collections. In *INTERACT '03*, pages 196–203, 2003.

[5] A. Girgensohn, J. Boreczky, P. Chiu, J. Doherty, J. Foote, G. Golovchinsky, S. Uchihashi, and L. Wilcox. A semi-automatic approach to home video editing. In *UIST '00*, pages 81–89, 2000.

[6] J. Heer, N. S. Good, A. Ramirez, M. Davis, and J. Mankoff. Presiding over accidents: System direction of human action. In *CHI '04*, pages 463–470, 2004.

[7] K. L. Kroeker. Photography's bright future. *Commun. ACM*, 52(2):11–13, 2009.

[8] M. Kumano, K. Uehara, and Y. Ariki. Online training-oriented video shooting navigation system based on real-time camerawork evaluation. In *ICME '06*, pages 1281–1284, 2006.

[9] R. Liss. Ten tips to better video. `http://www.camcorderinfo.com/content/ten_steps_better_video.htm`, 2004.

[10] S. K. Nayar. Computational cameras: Redefining the image. *Computer*, 39(8):30–38, 2006.

[11] Picasa. `http://picasa.google.com`, 2010.

[12] D. Ramachandran, J. Canny, P. Dutta Das, and E. Cutrell. Mobile-izing health workers in rural India. In *CHI '10*, pages 1889–1898, 2010.

[13] Shutterfly. `http://www.shutterfly.com`, 2010.

[14] V. Soo. Good video tips. `http://www.camcorderinfo.com/content/good_video_tips.htm`, 2001.

[15] R. H. Thaler and C. R. Sunstein. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press, 2008.

[16] W.-Q. Yan and M. S. Kankanhalli. Detection and removal of lighting & shaking artifacts in home videos. In *MULTIMEDIA '02*, pages 107–116, 2002.