Creating expository documents with web-based authoring and heads-up capture

Scott Carter, Pernilla Qvarfordt, Matthew Cooper
FX Palo Alto Laboratory
Palo Alto, CA 94304
{carter,pernilla,cooper}@fxpal.com

Ville Mäkelä
TAUCHI, School of Information Sciences
University of Tampere, Finland
Kanslerinrinne 1, FI-33014
ville.mi.makela@sis.uta.fi

ABSTRACT

Tutorials are one of the most fundamental means of conveying knowledge. Ideally, tutorials not only *describe* each step with text or via audio narration but *show* it as well using photos or animation. In most cases, tutorial authors capture media from handheld mobile devices to import into these documents, but increasingly they use wearable devices as well. Here we explore the use of head-mounted capture for authoring tutorials. We have developed a media-capture tool for Google Glass that requires minimal attention to the capture device and instead allows the author to focus on creating the tutorial content. We describe a study comparing standalone (camera on tripod) versus wearable (Google Glass) capture showing that tutorial authors prefer wearable capture devices, especially when recording activities involving larger objects in non-tabletop environments.

1. INTRODUCTION

Tools for capturing and authoring tutorials typically support a limited set of media and thus restrict expressiveness. Past work has shown that relying on a single media type is rarely the best way to convey expository content [3]. For certain tasks, video has been particularly helpful relative to static graphics [6, 12]. This is intuitive since some tasks involve gradual progressions that can be difficult to capture in static photos (e.g., fluffing egg whites). Other tasks can require multimedia feedback (e.g., playing musical instruments). Video can also help coordinate a set of steps into a cohesive sequence. For example, the act of kicking a football can be depicted by a series of static shots: lining up the foot, striking the ball at a particular spot, following through, etc. Without seeing these individual elements combined in a swift strike, it can be difficult to verify the correctness of the composite end result. Furthermore, using video does not preclude integrating static content – many video editing tools support the integration of static photos that can be "played" for some period of time within the video. Semi-automated systems can also help condense expository video into more consumable clips [2]. However, past work has also shown that video is not the best format for all learning tasks [11]. In some cases, the best approach involves combining text with static graphics [7].

For these reasons, mixing different types of media in the authoring process is critical for creating effective how-to videos (how-tos). Capturing content in the medium best suited to convey information is similarly important. Our early needfinding work indicated that many tutorial authors had difficulty recording complicated procedures with standard mounted-cameras, suggesting a role for head-mounted capture. This approach stands to benefit end users as well since past work has suggested that first-person video instruction can improve performance on assembly [9] and learning [10] tasks.

In this work, we explore the use of head-mounted capture for tutorial content creation. We built a set of tools for tutorial content creation, including a head-mounted content capture application. The how-to authoring environment allows authors to tailor the media capture, authoring, and annotation approach to the task being documented. Using these systems, we report the results of a user study focused on the use of head-mounted devices to compose how-to videos.

2. SHOWHOW: A TUTORIAL AUTHORING SYSTEM

In our previous needfinding work, we interviewed nine tutorial creators about their how-to creation practices and were able to observe two participants as they jointly created a how-to guide [1]. We found that:

- The capture process often is fundamentally subordinate to performing the documented activity itself. Authors are frustrated when their capture devices obstruct or complicate completing the activity. They are also likely to use familiar tools such as their camera phones, and are willing to trade capture quality for increased usability and convenience.
- Authors need to be able to choose a capture device suited to the context of the activity – what is being documented and where – to make media capture as unobtrusive as possible.
- Different tutorials have different capture requirements. Tabletop tutorials tend to focus on the construction of smaller objects and involve fewer camera-angle changes and longer shots. On the other hand, how-tos concerning other tasks (e.g., automotive, home repair, etc.) necessarily require a wider variety of camera angles and shorter duration shots.
- Authors had difficulty creating a narrative using video exclusively. Many content creators wanted to import and link to external content as well as bookmark and annotate important

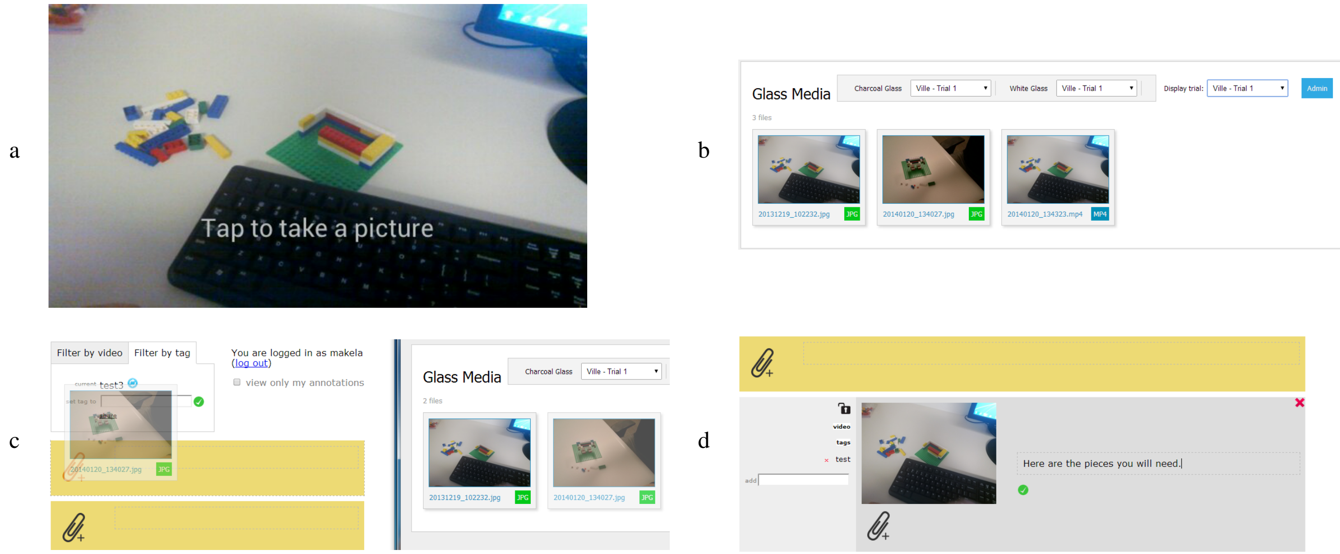


Figure 1: Capturing media with the ShowHow Glass application. Authors snap a photo or record a video with the Glass application (a). Media is automatically synchronized to the ShowHow web server and client (b). From there, authors can drag media into ShowHow tools (c). Next, the author can add text annotations describing the captured content (d).

sections of their video. They wanted to link dynamically to other examples or versions of the object or activity they were documenting.

The last item is consistent with [14] that found that how-to "sharing occurs within and across a collection of communication tools without any centralized control."

Based on these findings we built a web-based tool with support for authoring and annotating tutorials [1]. However, pilot studies revealed it to be frustrating to use. Ultimately we found that the interface failed to integrate into the flow of users' workaday tasks, largely because it forced users to interact with media indirectly. We then redesigned the system, ShowHow, to emphasize direct manipulation and re-mixing of multimedia content. Now, for example, users can drag media directly onto the page and our system will automatically create a well-formatted annotation. Importantly, the system supports representing, integrating, and manipulating a wide variety of media formats, including audio, PDF, photo, video, animated GIF, and text. Users can capture content with ShowHow capture applications or a third-party mobile client of their choosing, and use ShowHow's HTML5-based drag-and-drop authoring tools to assemble captured media into a tutorial.

2.1 Head-mounted capture

Head-mounted devices offer a potential opportunity to shift authors' attention and effort away from capture devices and towards completing the activity to be documented. Holding a mobile device can limit the user's range-of-motion, but even when the user places the device on a tripod they still must frame the content with respect to a viewfinder that may be difficult to see. Tripod mounts can also be tedious or difficult to move between and, especially, during shots.

For these reasons we decided to build a wearable capture application. Our goal was to build as unobtrusive a tool as possible. As previous needfinding work showed, it is critical that the wearable system does not force the user into unnatural actions or obscure the subject of the tutorial. At the same time, what the system records should be predictable by users. If the user is unsure whether their actions are within the captured field of view, they will exert effort to center the activity within the recording frame. For this reason, mounting the capture device to the head near the eyes is ideal.

A head-mounted system could be similar to augmented reality (AR) displays. However, unlike AR systems (e.g., [4] and [8]) which primarily support tutorial *access*, in a wearable *capture* system the user needs only straightforward methods to start and stop recording, review their recording, snap photos, etc. That said, simpler head-mounted capture devices, such as GoPro cameras [5], fail to provide enough interactive feedback for users. They may also require users to consult a mobile device during and after recording, which can be cumbersome.

In the end, we chose to use Google Glass since it sits in between these two extremes. Its design is relatively unobtrusive (e.g., the display is not a goggle-like wraparound display but rather a svelte form-factor that obscures only a fraction of the visual field, it does not require wires that could hinder movement during recording, etc.); it includes a display for simple feedback; it includes a camera near the eye; and it includes modest interaction methods (e.g., touch commands).

Our aim was to reduce required interaction during content capture and allow how-to authors to focus on their message rather than on the capture tool. However, the built-in camera application on Glass has several drawbacks. First, there is no preview for taking pictures, which forces users to take photos without knowing how well the content of interest is framed. Second, to record clips longer than ten seconds users must make a separate selection. Third, at the time none of the applications available supported uploading captures automatically – they all required the user to specifically select a photo or video and then "share" it with an application that would then upload the media to a remote service.

We developed our own capture application for Google Glass with which users can take pictures and record videos with glanceable preview. The application automatically uploads captured content to the ShowHow server and imposes no video duration restrictions (Figure 1). The Glass application is built on Android platform version 4.0.3. It uses the Glass Development Kit library's gesture detector to recognize different gestures on the Glass touchpad. A

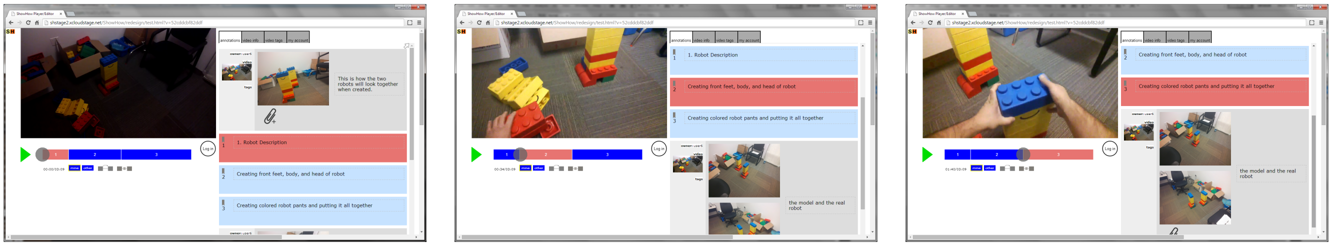


Figure 2: A video-based tutorial created by a participant in the study. This tutorial includes three three steps (bookmarks) as well as annotations at the beginning and end with photos showing the full figure.

separate component handles uploads to the server.

The application consists of three views. From the main menu, users select whether they want to take pictures or record videos. To navigate, the user swipes forward or backwards on the touchpad attached to Glass' frame and taps to select. Users return to the main menu by swiping down on the touchpad. In picture view a single tap takes a picture, while in video view the first tap starts and the second tap stops the recording (Figure 1a). Glass' display shows a viewfinder while in capture mode, making the recording glanceable. Media is uploaded automatically, and the upload progress is updated in the top-right corner of the screen of Google Glass. The Glass application uploads files asynchronously and they appear in the web application as they become available. After media are uploaded, they are visible in ShowHow (Figure 1b) in the order they became available. Now media can be dragged-and-dropped into a ShowHow tool for authoring (Figure 1c,d). In this way, authors can integrate captures from Glass into either the video-based or document-based tutorials described below.

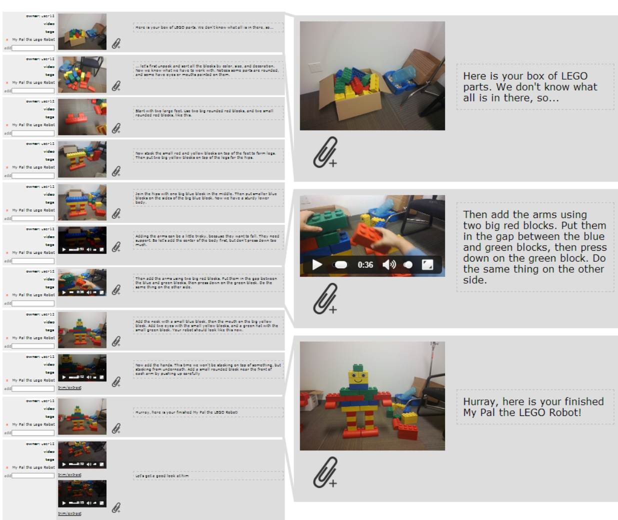


Figure 3: A document-based tutorial created by a participant in the study. This 11-step document uses a combination of photos, videos, and rich text to illustrate the construction of a robot using large blocks. The first, seventh, and tenth steps are shown in detail (right). Note that in the seventh step the video starts at 36 seconds, indicating that the user trimmed the original video clip.

2.2 Authoring how-tos with ShowHow

The head-mounted capture system is integrated with the authoring tool. ShowHow supports two separate methods for authoring: *video-based* and *document-based*. How-to content is commonly structured according to the *steps* that comprise the documented activity. Both authoring approaches allow users to combine different types of media to craft tutorials that optionally include step boundaries, step sequence, step importance, and references to other content or activity beyond the immediate scope of the current how-to. The key distinction is that the video-based approach fundamentally relies on a *temporal* organizational metaphor wherein the system *plays through* the video and associated bookmarks, while the document-based approach relies on a *spatial* organizational metaphor wherein the user *scrolls through* content.

The process of creating a *video-based* how-to involves 1) uploading raw content; 2) editing content; and 3) adding bookmarks and multimedia annotations to augment the how-to. To support rapidly combining and re-mixing short clips, we built an HTML5-based video creator. Authors can drag-and-drop video clips onto the tool, drag to reorder clips, and play each clip individually. They can also produce a higher quality clip with any external editing tool and upload it to our system to take advantage of our tools' rich multimedia annotation abilities. Figure 2 shows an example of a video-based tutorial created in our study. When the author is satisfied with the order and edits of individual clips she clicks a button to create a composite clip. Once processing is complete, users can view the video and begin adding markup.

The *document-based* approach lays out each step in the process separately using dedicated text and media. Figure 3 shows a document-based tutorial created with ShowHow. To start, a user drags media into the tool. This creates a "tag" where the user can add annotations. As in the video-based case, annotations include text and an arbitrary number of media elements. Each annotation is automatically assigned the current tag as it is created. The temporal sequencing of the video clips within the video-based tutorial corresponds to the vertical arrangement of the multimedia annotations in the document-based tutorial.

The document approach has some advantages over the video approach: steps can be skimmed with a glance, searched more directly, and never need to be "paused." In contrast, the video approach is more dynamic and doesn't require manually advancing step by step. We examine head-mounted capture in conjunction with authoring both types of how-tos.

3. USER STUDY

To assess the prospects of wearable capture for tutorial creation, we compared our head-mounted capture system with portable cameras on tripod in two common tutorial creation scenarios: working

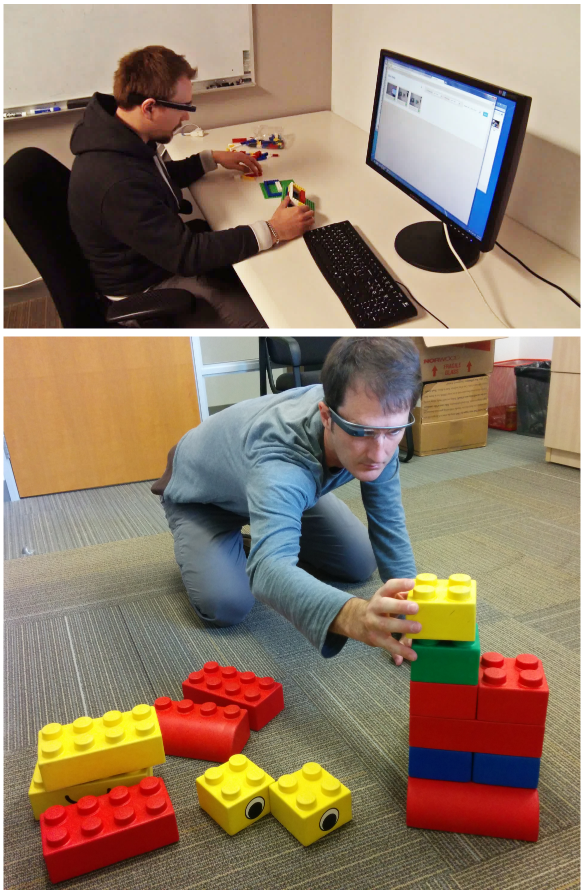


Figure 4: A user completing the table-top task while recording with Glass (top). Media is synchronized automatically to the web page shown on the monitor. For the floor task participants used foam blocks much larger than their standard Lego counterparts (bottom).

at a table and freely moving around a larger object. These two situations impose different requirements that influence how creators approach and experience the media capture process. A camera on tripod can require extensive manipulation which may limit the number of angles a person chooses for recording the objects of interest, however, it may produce higher quality media. Our study focuses on users' experiences with head-mounted capture and tripod-based capture in both of these settings.

As capture devices we used our Google Glass ShowHow capture application and a camera (Nexus 4) on tripod. The Nexus 4 was set up to automatically save media to a Dropbox folder which would upload the media to a computer for editing, simulating the automatic upload implemented in our Glass ShowHow capture application. In both conditions, the tutorial was authored using the ShowHow authoring tool by dragging media into the tool from either the ShowHow web client (Glass) or the Dropbox folder (Nexus 4).

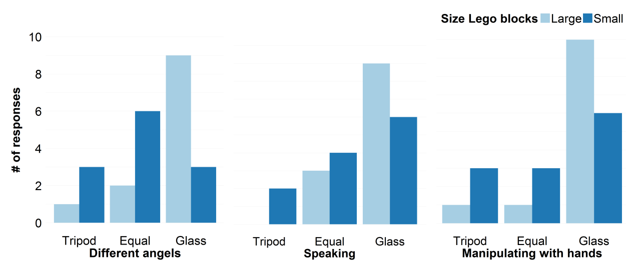
We asked twelve people to create two tutorials demonstrating how to build a robot with Lego. The two tutorials were created in separate sessions. The participants used different capture devices for the two tutorials. In one tutorial, the participants used regular size Lego blocks seated at a table (Figure 4, top), while in the other they used larger foam bricks on the floor (Figure 4, bottom). Before creating the tutorial, the participants were asked to design the Lego robot with the only restriction that it should be at least eight layers tall. During the design phase, participants were encouraged to try out the recording device and authoring tool. They were free to choose the structure and focus of their tutorial, and were asked to frame the video and photos to best suit their needs by moving the tripod or changing recording position while wearing Google Glass. The only requirement was that the final tutorial should be sufficient to enable someone to build an exact replica. After each of tutorial, participants answered questionnaires about their experience with our tools. The participants spent on average 23 min (SD=15.0) per study session.

Our authoring tools are created for amateur tutorial producers. Since part of our aim was to explore the ease of use of the tool-set, we used infrequent tutorial producers for our study. All participants had experience with video editing and making tutorials, although most (10) participants make tutorials less than once a year. The remaining two made up to six tutorials per year.

The order and combination of the three independent variables (activity location, capture device and authoring method) was balanced over all participants. All participants used both capture devices and authoring methods (video-based and document-based).

4. RESULTS AND DISCUSSION

Many factors influence the choice of capture devices and authoring tools. In our discussion, we focus on the creators' media capture experience and the quality and characteristics of the captured media.



(a) (b) (c)

Figure 5: Number of participants preferring Glass, neutral preference (equal) and Camera + Tripod (Tripod) when showing small vs. large model from (a) different angles, (b) showing a model and speaking and (c) showing a model and manipulating an item with their hands.

4.1 Experience with the capture devices

We asked study participants to compare the capture devices in activities with small versus large models (Figure 5). We found that participants significantly preferred Glass for capturing a large model, but report equal preference for either device with small models (Wilcoxon signed rank test: $V = 3.5, p < 0.05$). Also users preferred head-mounted capture while speaking and while working with their hands. Participants' comments reflected feeling comfortable with head-mounted capture when working with large objects or actively moving. Five of twelve participants felt that camera and tripod would be best suited for recording stable video and when no movement or different angles are required, such as detailed work in a static location.

For content creators, feeling confidence in the capture method is critical. Our data indicates that participants found head-mounted capture was straightforward to use, provided good feedback about what was in view, and required minimal setup. Specifically, participants appreciated that the camera would share their field of view while building the robots (mentioned by 6/12). The following are representative statements from two participants:

> Once I got the hang of it, I was able to check with a glance, the recorded video framing/quality etc. without taking my eyes of the model/task at hand (Participant 2)

> It's more natural and what you see is what you get so less time wasted in adjusting the camera or the object. (Participant 4)

Users also reported some difficulties using Glass to capture photos and videos. Other participants felt they had to regularly check that the camera was recording what they wanted (3/12). One participant said: "Occasionally this distracted from the task as I was looking at what the viewfinder was displaying" (Participant 8). Those that did not check the viewfinder during recording sometimes noticed later that objects they intended to show in the video were outside the frame either partially or completely. Although Glass' line of sight matches well with the wearer's, the participants (5/12) pointed out that this is not always sufficient: "Sometimes I forgot to keep my head pointed to my hands" (Participant 10). Also participants noted that when recording smaller objects, Glass had limited ability to zoom in (2/12) or show certain angles (2/12). Moving the head closer to the object may not feel natural or comfortable. One participant explained: "Since you're wearing Glass certain angle e.g. near the floor is hard to record" (Participant 4). We noticed one participant when trying to capture a Lego block close to the floor tilting his head to get a good close view. However, this did not result in a very good shot of the block in the video when the video image suddenly tilted.

Participants had mixed feelings about the camera and tripod set up. Four of twelve participants considered this set up convenient, while others focused on drawbacks such as the difficulty seeing what the camera was recording (5/12), the need for additional adjustment of the camera angle (4/12), and the constraint of working within the camera's field of view (3/12). Participants 8 and 2 noted issues with the camera-on-tripod set up:

> I wasn't sure if the object was in the field of view of the camera, and it was kind of complicated to check if the object is being viewed and documenting the build process at the same time. (Participant 8)

> The initial tripod setup was non-trivial and I was less inclined to mess with it once it was setup and so I ended moving the model around and closer to the camera to get different angles and better shots. (Participant 2)

According to our participants, the main advantages of using a camera on a tripod were good control of the camera angle (5/12) and the good quality of the resulting video (4/12).

Our results indicate that tutorial creators value the convenience and confidence Glass gives them over potential quality issues. This trade-off was particularly evident when working with a larger model that required movement over a larger area. Set up with Glass is minimal: a tutorial creator can simply move to a new position. Feedback of how well the content is framed is accessible with a glance. Video can become more shaky, but the value of unrestricted movements and instant feedback was clearly higher for our participants.

4.2 Captured media

To explore how participants used the tools available to them, we examined the media captured for the tutorials, and how participants manipulated the camera or the robot to show different perspectives. On average, participants captured 8.5 photos (SD = 10.80) with Glass and 7.0 photos (SD = 5.84) with camera + tripod, and 3.1 video snippets (SD = 3.68) with Glass and 4.5 video snippets (SD = 2.86) with camera + tripod. The total video duration was 144 s (SD = 79.7) for video captured with Glass and 148 s (SD = 143.5) for video captured by camera + tripod. None of these differences were significant. The location did not impact either the number of photos or video snippets or the captured video duration. The tutorial's topic dictates to some extent how much media the participants need to capture to convey their intended message. In this study, the topic is held constant so that we can examine whether different capture tools produce different strategies for capturing the necessary media.

An object can be depicted from different angles by either moving the camera or moving the object itself. When working with regular (smaller) sized Lego blocks, the participants rarely moved the camera (M = 0.4 times, SD = 0.79) and instead manipulated the Lego model to show multiple views. The participants were just as likely to move the Lego model when capturing with Glass as with camera + tripod, 4.6 moves per tutorial (SD = 6.29). Moving a large object is more difficult, and thus the participants moved the camera and tripod on average 4.6 times (SD = 5.53) with the larger bricks. Clearly, the need to frame the Lego robot from different perspectives was equally large when creating the tutorial at the tabletop as on the floor, only the method differed.

When focusing on shifts in framing (e.g. changes in angles, pan, zoom etc), we found that participants using Glass made significantly more framing shifts per minute (M = 6.0 shifts, SD = 4.78) within the recorded videos compared to using camera + tripod (M = 1.4, SD = 2.91; $F(1, 10) = 15.00$, $p < 0.01$). In fact, 69% (SD = 34.7) of the framing shifts in videos captured by Glass were made within a video snippet and 31% were made between video clips and/or photos. When capturing with camera + tripod, 33% (SD = 37.5) framing shifts were done within a video snippet. In addition, most of the shifts in framing were done while working with the large Lego bricks on the floor (Floor: M = 6.2 shifts per minute video snippet, SD = 5.02; Tabletop: M = 1.2, SD = 1.76; $F(1, 10) = 18.2$, $p < 0.01$). The large majority of the framing shifts seated at the table were done when the participants used Glass (M = 2.3, SD = 1.91). While capturing with Glass, the framing shifts were not always intentional, but a byproduct of participants' head movements.

On average, the participants retained 89% of the captured video in their final edited tutorials. The participants often trimmed the start and end point of video snippet as these generally contained more inadvertent framing shifts. Overall, these results show that participants using Glass to record larger objects more often changed the framing of the object to better convey important aspects of the building process compared to when camera + tripod was used. The reason may be that a change in framing can easily be done by moving the head or the body when wearing Glass. The threshold for doing a framing shift with a camera on tripod appears to be considerably higher.

4.3 Tutorial media quality

We asked three judges, professional and amateur photographers and videographers, to evaluate each tutorial's media quality. The media used in the tutorials (photo, video and audio) was rated on seven-point scale based on its sharpness, focus, framing, etc. Media quality is of course only one aspect of the tutorial's overall effectiveness. However, it is important because the media must illustrate important steps. If the video is too shaky, badly framed or the audio inaudible then a tutorial consumer will have more trouble following it. Of the ten top rated tutorials on media quality, seven were recorded with Glass and three with camera on tripod. In comparison, of the ten tutorials with the lowest ratings, four were recorded with Glass and six were recorded in the tripod condition.

These results are interesting since the participants perceived improved recording quality when using camera on a tripod. A head-mounted camera is more sensitive to head movements and the camera angle on Google Glass is not quite designed for recording manipulations with the hands close to the body. For this kind of recording, the participants needed to consciously frame the capture. However, through the complete authoring process, the participants managed to create tutorials with equal or better quality using head-mounted capture than using a camera on a tripod.

In sum, the study indicates that users find head mounted capture at least as effective as camera + tripod across our conditions, and for larger activities it is preferred. Working at a tabletop, camera on a tripod is an equally good alternative. The use of head mounted devices enables users to readily document their activities in a natural manner from multiple vantage points. The content they obtained in turn was judged to be of equal or higher quality.

5. CONCLUSION AND FUTURE WORK

Creating a tutorial requires capturing descriptive and procedural content and rendering it so that end users can comprehend and recreate the steps taken. In our study, we have seen that seamless capture of tutorial content allows the authors to focus on exposition. Head-mounted capture was perceived as easy to use, having a minimal set up with good feedback, and a first person view of the content. Our users preferred to capture tutorial content with our head-mounted capture application over a camera on a tripod, even when the camera on tripod included the same convenient features such as automatic upload. When handling smaller objects on a desktop, the camera on tripod and head-mounted capture were viewed to be equally useful. The Glass application was seen as particularly useful for showing tutorial content with large objects where authors needed to move around to show different viewpoints. Head-mounted capture was also preferred for adding natural narration.

Our findings support the idea that head-mounted devices that integrate capture and some form of real-time feedback, "could introduce a new form of interactive manual into the world" [13]. Once capture tools are paired with complementary access applications (which we are in the process of building), users will be able to view and navigate both video- and document-based tutorials on a head-mounted display and capture system. With this end-to-end system, we may come to use these devices to construct threaded conversations with family, friends, and colleagues about how to interact with everyday objects in our world.

Additionally we found that a close integration between capture and authoring tools streamlines the authoring process. We believe that our suite of web tools allows users to combine a variety of different media to convey tutorial material. We intend to evaluate how the different metaphors in the authoring model impact how and what tutorial creators capture. We believe it is important to understand the whole cycle of capture, authoring, and tutorial access to be able to design tools suited for the unique multimedia capture and consumption requirements of tutorial content.

6. REFERENCES

[1] Carter, S. Adcock, J., Cooper, M., and Branham, S., Tools to support expository video capture and access. *Education and Information Technologies Journal*. **19**(3): 637–654, 2013.

[2] Chi, P-Y, Liu, J., Linder, J., Dontcheva, M., Li, W., and Hartmann, B., DemoCut: Generating concise instructional videos for physical demonstrations. *Proc. ACM UIST*, 141–150. 2013.

[3] Feiner, S., McKeown, K. R. Automating the Generation of Coordinated Multimedia Explanations. *IEEE Computer*. **24**(10): 33–41, 1991.

[4] Feiner, S., MacIntyre, B., and Seligmann, D. Knowledge-based augmented reality. *Communications of the ACM*. **36**(7): 53–62, 1993.

[5] GoPro. http://gopro.com/.

[6] Grossman, T. and Fitzmaurice, G. ToolClips: An investigation of contextual video assistance for functionality understanding. *Proc. ACM CHI*, 1515–1524. 2010.

[7] Harrison, S. M. A comparison of still, animated, or nonillustrated on-line help with written or spoken instructions in a graphical user interface. *Proc. ACM CHI*, 82–89. 1995.

[8] Khuong, B. M., Kiyokawa, K., Miller, A., LaViola Jr., J. J., Mashita, T., and Takemura, H. The Effectiveness of an AR-based Context-Aware Assembly Support System in Object Assembly. *Proc. IEEE Virtual Reality*, 57–62. 2014.

[9] Kraut, R. E., Fussell, S. R., and Siegel, J., Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, **18**(1):13–49. 2003.

[10] Lindgren, R. Generating a learning stance through perspective-taking in a virtual environment. *Computers in Human Behavior*, **28**(4):1130–1139. 2012.

[11] Palmiter, S., Elkerton, J., and Baggett, P. Animated demonstrations vs. written instructions for learning procedural tasks: A preliminary investigation. *International Journal of Man-Machine Studies*, **34**(5):687-701. 1991.

[12] Pongnumkul, S., Dontcheva, M., Li, W., Wang, J., Bourdev, L., Avidan, S., and Cohen, M. F. Pause-and-play: Automatically linking screencast video tutorials with applications. *Proc. ACM UIST*, 135–144. 2011.

[13] Svenvold, M. The Disappearance of the Instruction Manual. *Popular Science*, February, 2015. http://www.popsci.com/instructions-not-included/.

[14] Torrey, C., McDonald, D., Schilit, W., and Bly, S., HowTo pages: Informal systems of expertise sharing. *Proc. ECSCW*, 391–410. 2007.

7. AUTHOR BIOS

Scott Carter (PhD) is a senior research scientist at FX Palo Alto Laboratory. His primary research focus is developing innovative multimedia user interfaces. You can contact him at carter@fxpal.com

Pernilla Qvarfordt (PhD) is a senior research scientist at FX Palo Alto Laboratory. Her research is focused on understanding human attention and behavior when interacting with multimedia environments. You can contact her at pernilla@fxpal.com.

Ville Mäkelä is PhD student in the School of Information Sciences at the University of Tampere in Finland. You can contact him at ville.mi.makela@sis.uta.fi.

Matthew Cooper (DSc) is a senior research scientist at FX Palo Alto Laboratory, leading the Interactive Media group. He develops content-analysis techniques to support interactive multimedia applications. You can contact him at cooper@fxpal.com.