ToolScape: Enhancing the Learning Experience of How-to Videos

Juho Kim
MIT CSAIL
Cambridge, MA 02139 USA
juhokim@mit.edu

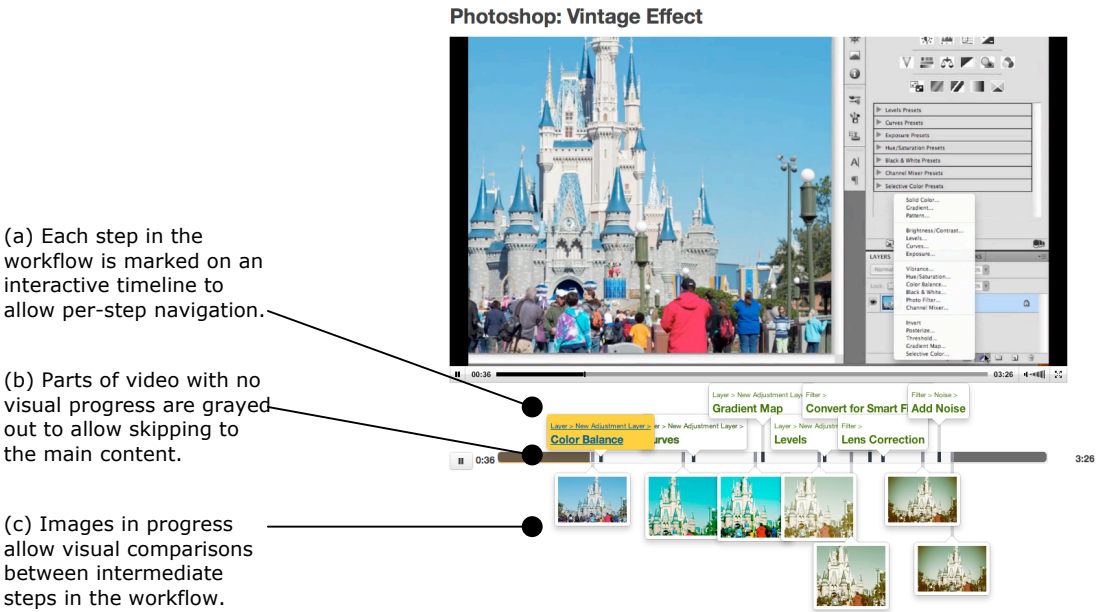


Figure 1. ToolScape gives a learner control when watching a how-to video with various non-sequential ways to navigate a workflow.

CHI 2013 Extended Abstracts, April 27–May 2, 2013, Paris, France.
ACM 978-1-4503-1952-2/13/04.

Abstract
Video tutorials on the web have gained popularity in various domains, but most video repositories are not designed to support the unique content and structure of how-to videos. Learners face difficulty in finding relevant videos and applying the skills embedded in a video clip. We introduce ToolScape, a video browsing interface with a storyboard summarization and an interactive timeline. It allows learners to quickly scan, filter, and review multiple videos without having to play them. Learners can also jump to or repeat a particular step within a clip by clicking interactive indices on the timeline. In a within-subjects study where participants engaged in end-to-end design tasks with ToolScape and a control interface based on YouTube, the participants using ToolScape rated their design work higher and showed a higher gain in self-efficacy. External raters ranked designs using ToolScape higher.

Author Keywords
Video tutorials; How-to videos; Video interface; Learning support.

ACM Classification Keywords
H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems.

Introduction

Video tutorials on the web have expanded the amount and diversity of learning options available, affecting the way creative workers find, access, and learn from videos. These how-to videos span a variety of domains, including complex software applications, cooking, makeup, craft, and art. Searching on YouTube for a typical task in graphical editing software Photoshop such as "*removing an object in Photoshop*" returns over 4,000 video results. But few of them may be relevant to a particular learner because videos vary in style, quality, skill level, and context. The poor information scent of video tutorials makes it difficult for learners to *find* videos and *apply* skills catered to their tasks or learning goals.

Learners often start from a search interface to find a tutorial that is useful or relevant. Search interfaces provide metadata surrogates [1] that help learners assess the relevance of video search results. Unfortunately video surrogates are typically limited to title, view counts, and thumbnails, and do not directly incorporate metadata about the procedure itself, like tool use, workflow ordering, or required skill level, that learners can benefit from.

Different challenges arise when learners watch and try to apply skills from a video tutorial. How-to videos contain multiple steps to reach a goal, and it is common for learners to follow along step-by-step or refer to a specific step in the workflow. Because comprehensive and accurate indices are missing, non-sequential access may become frustrating for learners. They rely on thumbnail previews and imprecise estimates to navigate between steps.

This research aims to enhance the learning experience when browsing and watching how-to videos. We hypothesize that video summarization methods and interaction techniques customized to how-to videos can improve learner satisfaction and performance. This research focuses on video tutorials on graphical design software, specifically Photoshop, due to its high penetration and the availability of large tutorial video repositories on the web.

Related Work

This paper builds on the active body of research on enhancing the tutorial experience. Systems automatically generate interactive tutorials by demonstration [3, 6] or help learners to follow along instructions [11].

Previous research has also looked at the value of mixed format instructions. Clark and Mayer [4] note that animations are good for physical procedures, while still images are good for conceptual processes. A how-to workflow often involves both types of processes. For example, in Photoshop, planning the overall design approach might be conceptual, but using selection tools to select an irregular object might be physical. We argue that video interfaces can benefit from incorporating more images and text.

Kong *et al.* [9] report that text+image is preferred to text only or graphical only instructions. Chi *et al.* [3] show that learners using a mixed tutorial (static+video) made fewer errors than using static or video alone. Their work attempts to incorporate video clips into a step-by-step tutorial. Our work is indeed an attempt at the inverse: can we integrate the step-by-step nature into a video browsing and watching framework?

Annotating Video Workflows

What kind of information from video tutorials, then, should be extracted and displayed? A design opportunity for enhancing how-to videos is that they have a more defined structure than most other videos. First, tasks are visual in nature, and progress can be visually tracked. Capturing intermediate works in progress and displaying them can help learners make sense of a workflow. Second, a set of actions or tools identifies a step from one version to next. A list of used tools helps learners comprehend how an effect is accomplished. We claim that annotations for how-to videos should combine the two properties to accurately summarize an entire workflow, therefore collecting both works in progress and tools between steps.

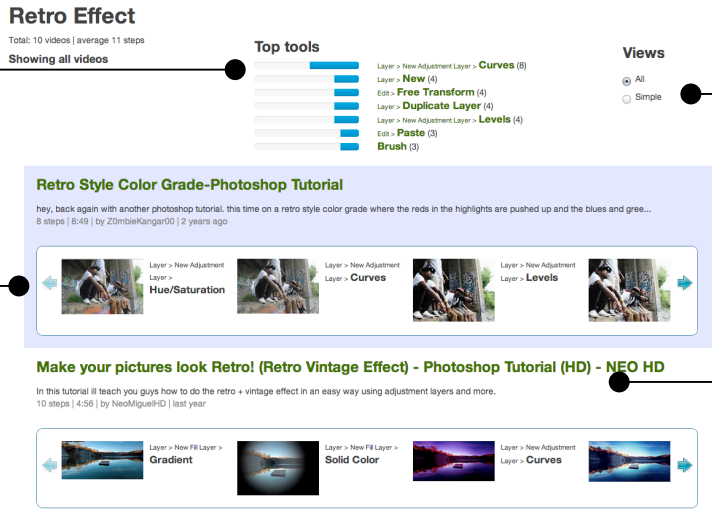
System Design

ToolScape is a web-based interface for browsing and watching how-to videos. Powered by annotations of a video workflow, namely commands and work-in-progress images, ToolScape provides a browsing interface with the Storyboard summarization and faceted search, and a player interface with an interactive timeline. Both interfaces are built with HTML5, CSS3, JavaScript, and an open-source video player. We followed an iterative design process with multiple rounds of pilot user feedback and refinement.

ToolScape player (Figure 1) allows learners to easily jump to or repeat a particular step inside a video clip without having to manually navigate a video player timeline slider. We use an interactive timeline to play annotated video clips as in existing systems [7, 11]. The top (Figure 1(a)) and bottom (Figure 1(c)) streams represent commands and works in progress, respectively. The visual separation allows scanning just the command names or works in progress.

Top tools show most frequently used tools in videos covering the currently searched effect (retro effect in this figure). Faceted navigation displays only workflows that include selected tools. Clicking a tool adds a filter, and multiple filters can be applied for more fine-grained filtering.

The Storyboard video summarization method lists keyframes. Keyframes include each step in the workflow (image) and a means to reach a step from the previous one (command).



Two view modes: *All* reveals the Storyboard summarization and *Simple* opens only before and after images.

Metadata display follows that of conventional video search interfaces such as YouTube or Vimeo. Title, description, length, uploaded date, and uploader information is displayed. The number of steps is calculated from the workflow and displayed in addition, to hint at the difficulty level of the video.

Figure 2. ToolScape makes browsing multiple how-to videos easier with the Storyboard summarization, faceted navigation, and filtering.

Self-rating & Self-efficacy

Education research shows that self-efficacy is an effective predictor of motivation and learning [2]. Motivation is especially important for aspirational learners outside of classroom. Positive self-assessment has also been shown to accurately predict learning gains [12].

Self-efficacy questions

The questions were adopted and modified from Dow *et al.* [5] to fit with the study context. In a scale of 1 (not confident at all) to 7 (very confident), the questions asked "How confident are you...":
- with solving graphic design problems?
- at understanding graphic design problems?
- with applying design skills in practice?
- with incorporating skills from video tutorials in your design?

A distinct feature in ToolScape player is the visualization of parts with visual progress (Figure 1(b)). Pilot user observations suggest that learners often want to skip unnecessary parts. The beginning and end of a video often include setup instructions *(e.g., opening Photoshop)* or personal comments *(e.g., advertising to rate the clip)*, and our annotations enable skipping to the point where the first command was issued. For the videos in our samples, 13.7% in the beginning and 9.9% at the end on average was time with no visual progress. This suggests that a user can save at least 20% of their watching time.

The browsing interface (Figure 2) allows learners to quickly scan, filter, and review multiple videos without having to play them. It displays a sequential workflow for each video using the Storyboard summarization method, which horizontally lists tools and before / after images for each tool. The summary generator samples and displays all frames specified in the given image and command annotations. To highlight the semantic difference between image and command, the summarization displays commands as text. This image+text representation visually distinguishes the two types of information, and further enables textual indexing and filtering with commands.

User Study

We conducted a laboratory user study to see if ToolScape helps users learn and apply new skills in a Photoshop design task. The study compared the skill-learning experience of ToolScape against a standard video interface, using the measures of self-efficacy, learner satisfaction, and performance.

The baseline interface has browsing and playing interfaces similar to YouTube. Its browsing page does not include the Storyboard summary, tool filtering, and view modes. It has a thumbnail for each video, along with basic metadata such as title, description, and length as can be found on YouTube's search results page. The playing page simply contains a video player.

We hypothesize the following:

H1 Learners with ToolScape complete design tasks with higher self-efficacy.

H2 Learners with ToolScape rate their work higher.

H3 Learners with ToolScape produce higher quality designs.

We recruited 12 novice Photoshop users (8 male) with a university mailing list and online community posting. The study was a within-subject design, with interface, task, and order counterbalanced. Each participant had two image manipulation tasks in Photoshop, which were to apply a retro effect and to transform a photo to look like a sketch. Baseline or ToolScape was the only allowed external help resource.

After a tutorial on the interface, the participant answers self-efficacy questions. Then a 20-minute task starts, and the participant can freely browse the given 10 videos and work on their task in Photoshop. After the task, the participant answers questions on task difficulty, self-rating, and interface satisfaction. We ask the self-efficacy questions again to observe changes. We compare self-efficacy gains between the interface conditions. The participant also scores each interface feature in the scale of how much it helped him or her during the task (1-not helpful at all, 7-very helpful).

After the study, four external raters evaluated the quality of the participants' designs. They ranked (1-

H1, higher self-efficacy for ToolScape, is supported. The mean gain in ratings for ToolScape and Baseline were 1.4 and 0.1, respectively. Mann-Whitney's U test on the difference in self-efficacy questions in 7-Likert scale shows a significant effect of interface *(Z=2.0586, p<0.05)*.

H2, higher self-rating for ToolScape, is supported. The mean ratings for ToolScape and Baseline were 5.3 and 3.5, respectively. Mann-Whitney's U test shows a significant effect of interface *(Z=2.6966, p<0.01)*.

H3, higher external rating for ToolScape, is supported. The rankings show high inter-rater reliability (*Krippendorff α=0.753*) for ordinal data. The mean rankings (lower is better) for ToolScape and Baseline were 5.7 and 7.3, respectively. A Wilcoxon Signed-rank test shows a significant effect of interface *(W=317, Z=-2.79, p<0.01, r=0.29)*.

best, 12-worst) the submissions based on how well the designs accomplish the given task. This rating method encourages a direct comparison between the designs and reduces individual variance in the ratings.

User Study Results

ToolScape had a positive effect on learners' belief in their graphical design skills (H1). Learners showed a higher self-efficacy gain with ToolScape. Participants rated their own work quality higher when using ToolScape (H2). External ratings suggest that they produced better designs with ToolScape (H3).

Non-sequential access and learner control of the playback were highly used and preferred. Participants clicked interactive indices on the timeline 8.8 times on average (σ=6.4) per task. Table 1 summarizes feature preferences. Most features of the player interface were highly rated. Users found the graying out of non-crucial regions to be very useful (6.5). Along with clicking to jump to images and commands (6.4), it suggests that supporting non-sequential access to keyframes is important. Participants noted, "*It was also easier to go back to parts I missed.*" (P4), "*I know what to expect to get to the final result.*" (P2), and "*It is great for skipping straight to relevant portions of the tutorial.*" (P1) We interpret that more control in navigating workflows allows learners to focus more on the task itself. This positive experience might have increased self-efficacy, which in turn might have promoted learning Photoshop skills better.

It is interesting to note that video length was less important metadata in ToolScape (3.5) than in Baseline (5.2). Mann-Whitney's U test shows a significant effect of interface *(Z=-2.6028, p<0.01)*. The reason might be that participants using ToolScape had more visual and direct cues to rely on for relevance evaluation than video length.

Top tools (4.7), tool filtering (4.6), and the number of steps (3.9) were the lowest rated features among those only in ToolScape. The result is not surprising because our database displayed only 10 videos at once, and the top tools or filter results did not provide much benefit. Top tools based simply on frequency is problematic because in many cases top-ranked tools are generic ones such as *New Layer* or *Duplicate Layer*. In the next iteration we plan to apply an algorithm such as TF-IDF to emphasize tools unique to the current task.

We additionally looked at time to completion, but it had no difference between interfaces. There are conflicting factors in play: the ability to skip unnecessary parts and higher accuracy in finding a specific moment in the video might shorten the task time, but users tend to have more confidence when opening a video due to improved information scent, which might lead to more exploration of workflows and videos.

Several users were concerned with information overload in the interactive timeline. The current timeline display suffers from occlusion when there are multiple adjacent short steps. Future iterations will strengthen favored features and address user concerns.

Large-Scale Annotation

In order for ToolScape to be of practical use, it is essential to collect annotations for a large number of how-to videos in an efficient and scalable way. An alternative approach would be to collect application context information at the tutorial recording time. But this approach might not scale, and ideally ToolScape can operate on top of *existing* videos readily available

Feature	TS	BL
After image	**6.7**	-
Visualizing non-essential, non-visual parts	**6.5**	-
Interactive indices for images and tools	**6.4**	-
Interactive timeline	**6.3**	-
Work-in-progress images in Storyboard	**6.1**	-
Thumbnail	-	**5.8**
Storyboard summarization	**5.7**	-
Before image	**5.6**	-
List of tools in Storyboard	**5.2**	-
Video title	5	**5.6**
Top tools display	**4.7**	-
Player thumbnail preview	**4.6**	4.3
Tool filtering	**4.6**	-
Number of steps	**3.9**	-
Video length	3.5	**5.2**
Video description	2.7	**3.8**
Upload date	**2.3**	**2.3**
Uploader	**2.3**	1.8

Table 1. Interface feature preference of ToolScape (TS) and Baseline (BL) sorted by 7-scale Likert score. ToolScape features, especially the ones providing non-sequential access to the workflow, were rated higher. Blank cells mean features absent in interface.

on the web. We are exploring different methods for annotating how-to videos after the fact.

Computer vision is a cost-effective and automatic way to collect annotations, but it requires high-resolution images for high accuracy and training data to yield good results. **Crowdsourcing** can be a viable solution to complement computer vision by providing low-cost training data. We are currently experimenting with alternative task designs to reach high accuracy [10]. As with other crowdsourcing systems, quality control and an associated cost rise remain a challenge. Our future work will mainly focus on **learnersourcing**, which leverages learners' activities as useful input to the system [8]. Learners are a motivated and qualified crowd who are willing to watch how-to videos for their learning purposes. We plan to inject quizzes while watching videos in ToolScape, whose answers serve as annotations and training data. We believe this mixed-initiative approach can produce additional learning benefits with well-designed quizzes and collect high quality annotations with low cost at the same time.

Acknowledgements

I thank Phu Nguyen for his help with the user study and crowdsourcing workflow design, and Joel Brandt and Mira Dontcheva for initial discussions, and Krzysztof Z. Gajos and Robert C. Miller for their advice and feedback. This work is supported in part by Adobe, and by Quanta Computer as part of the T-Party project. Opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the sponsors.

References

[1] Balatsoukas, P., Morris, A., and O'Brien, A. An evaluation framework of user interaction with metadata surrogates. J. Inf. Sci. 35, 3 (June 2009), 321–339.

[2] Bandura, A. Self-efficacy: toward a unifying theory of behavioral change. Psychological review 84, 2 (1977), 191.

[3] Chi, P.-Y., Ahn, S., Ren, A., Dontcheva, M., Li, W. & Hartmann, B. Mixt: Automatic generation of step-by-step mixed media tutorial. In Proc. UIST 2012, 93-102.

[4] Clark, R., and Mayer, R. E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. Wiley & Sons, 2007.

[5] Dow, S. P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D. L., & Klemmer, S. R. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. ACM TOCHI 17, 4 (Dec. 2010).

[6] Grabler, F., Agrawala, M., Li, W., Dontcheva, M. & Igarashi, T. Generating photo manipulation tutorials by demonstration. In Proc. SIGGRAPH 2009, 1–9.

[7] Grossman, T., Matejka, J. & Fitzmaurice, G. Chronicle: capture, exploration, and playback of document workflow histories. In Proc. UIST 2010.

[8] Kim, J., Miller, R. C., & Gajos, K. Z. Learnersourcing Subgoal Labeling to Support Learning from How-to Videos. In Proc. CHI 2013 EA, to appear.

[9] Kong, N., Grossman, T., Hartmann, B., Agrawala, M. & Fitzmaurice, G. Delta: a tool for representing and comparing workflows. In Proc. CHI 2012, 1027–1036.

[10] Nguyen, P., Kim, J., & Miller, R. C. Generating Annotations for How-to Videos Using Crowdsourcing. In. Proc. CHI 2013 EA, to appear.

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

[12] Schunk, D. Goal setting and self-efficacy during self-regulated learning. Educational psychologist 25, 1 (1990), 71–86.