Learnersourcing: Improving video learning with collective learner activity

by

Juho Kim

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2014

Author . . . .
Department of Electrical Engineering and Computer Science
May 16, 2014

Certified by . . . .
Robert C. Miller
Professor
Thesis Supervisor

Accepted by . . . .

Abstract

Millions of learners today use educational videos to master skills and take classes from online platforms such as YouTube, Coursera, or edX. However, when watching a video, learners face difficulty in accessing parts they want, understanding the overall structure, and seeing how other students learn. To address these issues, this thesis investigates data-driven video learning interfaces: interfaces that embed collective learner activity to improve video learning. We introduce *learnersourcing*, a general framework for motivating video learning activity and feeding the collected data back into the video interface. We present two learnersourcing methods. In passive learnersourcing, learners' natural interactions with learning material are automatically captured and analyzed to identify interaction patterns. In active learnersourcing, learners are prompted to provide input that is both pedagogically meaningful to learners and useful to the system. We are building prototype video learning systems to explore the design space of learnersourcing applications. For example, learners' second-by-second video interaction data from lecture videos improve video navigation, paid crowd workers extract step-by-step structure from how-to videos, and learners summarize individual steps from how-to videos into higher-level subgoals. With this conceptual framework for incorporating collective learning data into the learning experience, we are addressing technical and motivational challenges in building large-scale, video-based online education platforms.

Chapter 1

Introduction

Millions of learners today are using educational videos to master skills and take classes from online platforms such as YouTube, Coursera, or edX. While more attention has been given to creating more videos and providing open access to them, relatively less attention has been given to improving the experience of learning from these videos. Most existing video interfaces are not tailored to support learning. Possibly due to limitations in video interaction, delivery, and presentation, many learners resort to watching passively and linearly. While many learning theories show the value of interactive and constructive video learning [11, 7, 38, 37, 33, 34], it is challenging to implement the theories in open online platforms due to their large scale. Enabling techniques supported by the theories often require an extensive amount of customization, expertise, or manual effort.

To address this challenge, this thesis explores the feasibility of leveraging a massive amount of data generated by learners who interact with the videos. Data from natural and prompted learning activities provides an unprecedented opportunity to enable more interactive and constructive video learning. This thesis introduces *learnersourcing*, a set of data-driven methods and interaction techniques for improving learning, which are powered by data from collective learning activity. While the concept can potentially be applied to other learning materials, this thesis investigates learnersourcing in the context of instructional videos. The findings reported in this thesis have implications for applying learning theories to web-scale open learning environments, beyond small, in-person classrooms. This thesis

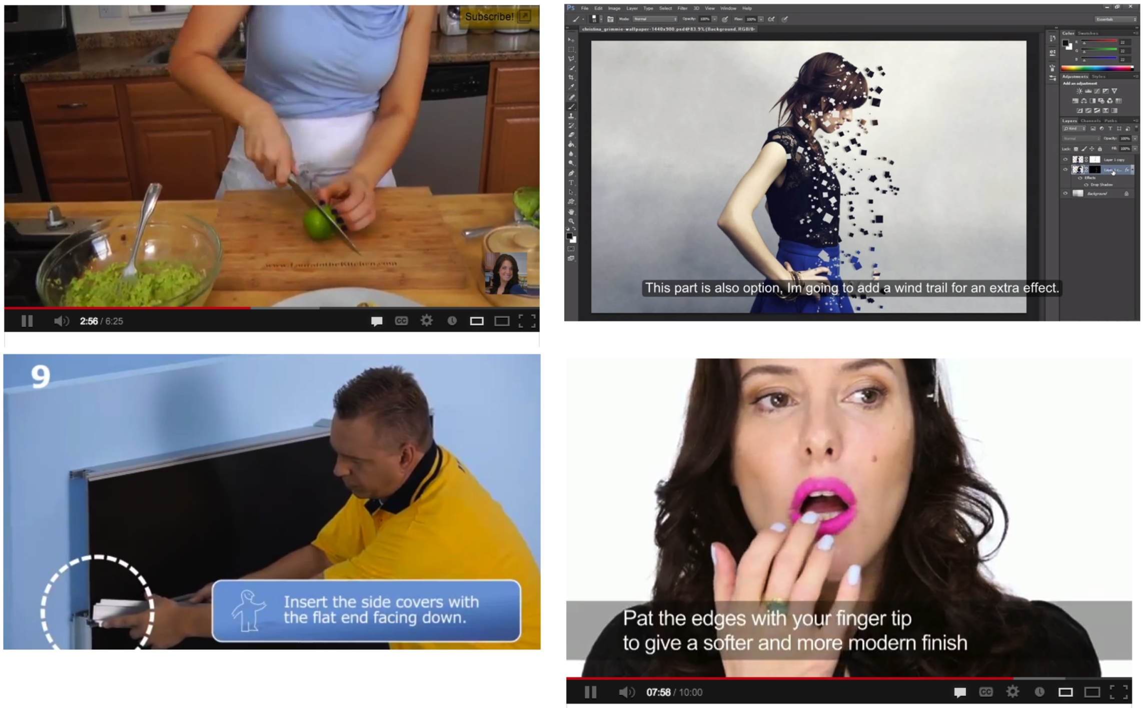

Figure 1-1: How-to videos include step-by-step instructions, spanning various domains including cooking, graphical design, home improvement, and applying makeup.

covers two types of educational videos, namely how-to tutorial videos and lecture videos.

How-to videos contain procedural instructions for how to complete a task in a step-by-step manner. A preliminary study shows the navigational and learning benefits of having step-by-step information about the solution, which was evaluated using ToolScape, a novel video player that displays step labels on the video timeline. To power ToolScape with the required data, this thesis introduces two methods for collecting step-by-step information from existing videos. The first method uses paid crowd workers on Mechanical Turk to extract low-level steps. The second method prompts learners who are watching the video to summarize what they learned, whose inputs are then processed by the system to extract high-level summaries of steps. When combined, the two methods can fully extract a hierarchical solution structure from how-to videos, by asking crowd workers to generate low-level steps and learners to group related steps into a higher-level goal. Finally, the thesis will report results from a user study designed to evaluate motivational, navigational, and learning benefits of having such solution structure handles generated by crowds and learners.

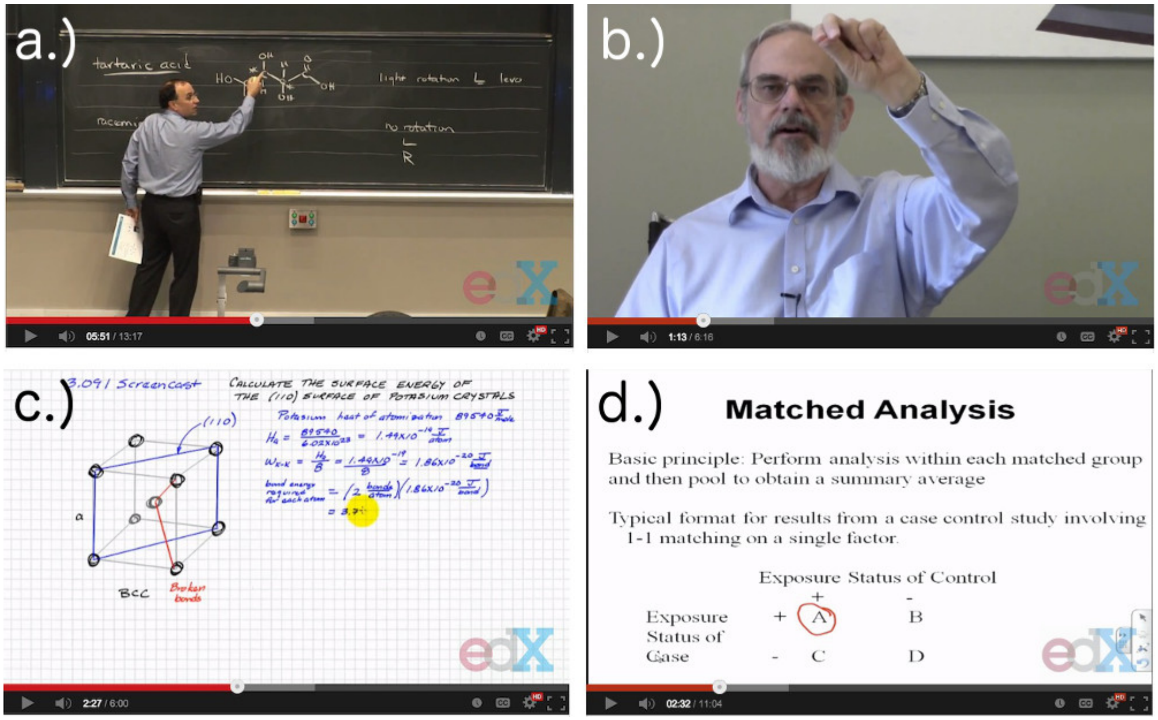

Figure 1-2: Lecture videos on the web show different styles, including a.) classroom lecture, b.) "talking head" shot of an instructor at a desk, c.) digital tablet drawing format popularized by Khan Academy, and d.) PowerPoint slide presentations.

Lecture videos can be commonly found in massive open online courses (MOOCs) from platforms such as Coursera, edX, or Udacity. An exploratory data analysis identified video interaction patterns by analyzing learners' second-by-second video interaction traces from four live MOOCs. The findings guided the design of LectureScape, a novel video player that leverages collective learner interaction data. LectureScape explores the design space of data-driven interaction techniques for educational video navigation. It introduces a set of interaction techniques that augment existing video interface widgets, including: a 2D video timeline with an embedded visualization of collective navigation traces; dynamic and non-linear timeline scrubbing; data-enhanced transcript search and keyword summary; automatic display of relevant still frames next to the video; and a visual summary representing points with high learner activity. Finally, the thesis will report results from a user study designed to evaluate motivational, navigation, and learning benefits of interaction techniques powered by learners' data.

Two primary technical contributions of learnersourcing are 1) methods for collecting and processing large-scale data from learners, and 2) data-driven interaction techniques to augment existing video players. The contributions are made possible by uniquely extending and combining the following three families of techniques: 1) crowdsourcing to collect and handle a large amount of learner input, 2) social computing to incentivize learners to participate and collaborate, and 3) content-based video analysis techniques such as computer vision and natural language processing to complement learner input.

Thesis statement: In large-scale video learning environments, learnersourcing encourages learner participation, enhances video content, and leads to better learning.

Chapter 2

Conceptual Framework

Learnersourcing is a set of techniques for collecting, processing, and presenting collective learning activity in online learning environments. The goal of learnersourcing is to improve the learning experience at scale with data generated by learners and their interactions with instructional materials and each other. With learnersourcing, we envision creating a data-driven learning experience that was not possible before. This thesis aims to present a conceptual framework for learnersourcing, design principles, approaches to address technical, pedagogical, and motivational challenges, and end-to-end examples.

Inspired by the concept of crowdsourcing, learnersourcing also attempts to collect small-scale contributions from a large group of users. A fundamental difference is that the crowd in learnersourcing is learners, who have the intrinsic and extrinsic motivation to learn new concepts and skills. This difference leads to a unique set of design considerations, in terms of incentive design, quality control, and task scope, and task difficulty. A crucial element in learnersourcing is to design learner's activity that is pedagogically useful, while the system collects useful information as a byproduct of such activity.

Depending on how learner input is collected, we define two types of learnersourcing: passive learnersourcing and active learnersourcing. **Passive learnersourcing** uses data generated by learners' natural interaction with the learning platform. Examples includes playing and pausing events from the video player, and browsing patterns between learning modules. On the other hand, **active learnersourcing** prompts learners to provide specific information with the purpose of collecting that data. Examples include asking learners to

Domain	Theory	Method	Properties	Impact
How-to videos	- procedural instruction - interactivity **- subgoal labeling**	- crowdsourcing - active learnersourcing with prompts	- mining solution structure by reverse-engineerin g a video	- crowd annotation method **- learnersourcing method + live deployment**
MOOC videos	- interactivity **- social (peer) learning** **- concept map** **- self-explanation**	- sync. discussion - passive learnersourcing with interaction traces **- active learnersourcing with prompts**	- large-scale data analysis - learner interest modeling - data-driven video interaction techniques	**- live deployment**
Office Mix videos	**- multimedia learning principles** **- social learning**	**- learnersourcing**	**- authoring tool intervention**	**- impact product**

Figure 2-1: Conceptual framework: each video domain is described with a summary of related theory, method, properties, and impact. Bold items indicate ongoing or proposed work to be completed before graduation.

summarize video segments they watched, to answer embedded questions, and to explain concepts discussed in the instructional material in their own words. In both cases, the system processes the collected data to further analyze how learners are using the instructional materials and improve future learners' learning experience.

With the presented conceptual framework in mind, I've structured my work around the following categories and summarized in 2-1:

- Domain: type of educational video
- Theory: learning theory our systems attempt to implement
- Method: technological approach to realize theory at scale
- Properties: notable characteristics of each approach
- Impact: Deployment (evaluation + impact): open source, real users, etc. publications, systems, intellectual contributions, experiments, evaluations, case studies

Chapter 3

Related Work

This thesis draws on three bodies of previous research: 1) designing interactive video instructions to improve the navigation and learning experience, 2) using collective user data to model and present user interest, and 3) crowdsourcing and communitysourcing techniques to motivate small-scale contributions at large-scale.

3.1 Video Navigation and Learning

3.1.1 Video Learning Theories

In designing user interfaces for instructional videos, higher interactivity with the content has been shown to aid learning [12, 46]. Tversky et al. [46] state that "stopping, starting and replaying an animation can allow reinspection", which in turn can mitigate challenges in perception and comprehension, and further facilitate learning. Semantic indices and random access have been shown to be valuable in video navigation [50, 32, 1] and the lack of interactivity has been deemed a major problem with instructional videos [18]. This thesis introduces user interfaces for giving learners more interactivity in video navigation, and learnersourcing methods for acquiring metadata handles to create such user interfaces at scale.

3.1.2 Video Navigation Support

Existing video summarization techniques use video content analysis to extract keyframes [3, 14], shot boundaries [32], and visual saliency [21]. To provide an overview of the entire clip at a glance and support rapid navigation, recent research has used a grid layout to display pre-cached thumbnails [36], short snippets [23] in a single clip, or personal watching history for multiple clips [2]. For educational lecture videos, Panopticon [23] has been shown to shorten task completion time in seeking information inside videos [41]. This thesis introduces alternative summarization techniques that use data from learners' interaction with the video. This method can be combined with content analysis to incorporate social signals in extracting highlights and navigational cues.

Another thread of research introduced techniques to support navigation of how-to videos, a sub-genre of educational video that includes procedural, step-by-step instructions about completing a specific task. Recent systems create interactive tutorials by either automatically generating them by demonstration [8, 15, 13], connecting to examples [43, 31, 13], or enhancing the tutorial format with annotated information [8, 9, 28, 30]. This thesis contributes crowdsourcing and learnersourcing workflows, which can provide annotations required to create these interfaces and further enable new ways to learn from tutorials.

3.2 Modeling Collective User Data

There is a rich thread of research in using collective user interaction history data to analyze usage patterns and improve users' task performance. Interaction history data is automatically collected by applications during normal usage. Examples include Web browsers logging Web page visit history, search engines capturing query history, and video players storing video interaction clickstreams such as play and pause events. Read Wear [20] presented a visionary idea in this space to visualize users' read and edit history data in the scrollbar. Chronicle [16] captured and provided playback for rich, contextual user interaction history inside a graphical application. Dirty Desktops [22] applied magnetic forces to each interaction trace, which improved target selection for commonly used widgets. Patina [35] separated individual and collective history and added overlays on top of the

GUI, to help people find commonly used menu items and discover new ways of completing desktop-related tasks. Causality [40] introduced an application-independent conceptual model for working with interaction history. This thesis introduces passive learnersourcing, a technique for using video interaction history to support common navigation tasks in video-based learning.

To model user interest in video watching, researchers have proposed features such as viewership [45], scrubbing [49], zooming and panning [5], and replaying and skipping [10] activities. SocialSkip [10] applied signal processing to replaying activity data in order to infer interesting video segments. Other work has used more explicit input from video watchers, including user ratings [42], annotations [45], and the "this part is important" button [44]. Most existing approaches introduce a modeling technique or data visualization. We take this data further to build new interaction techniques for video navigation, which prior work has not done. Also, we extend prior work on providing social navigation for lecture videos [39] to support diverse learning tasks.

3.3 Crowdsourcing and Communitysourcing

Learnersourcing is inspired by prior work on human computation [48], in which the system solicits human input for certain parts of the computation. Games with a Purpose [47] present computationally difficult problems (e.g., tagging objects from an image) as a game. Users simply play the game, and the byproduct of the gameplay is collected and processed by the system. What makes this approach scalable and motivational is that users do not have to know how their input is used by the system, because they have the strong motivation to play the game anyway. In learnersourcing, learners' tasks need to be designed so that they are intrinsically useful and enjoyable. Sometimes, however, knowing the bigger cause might additionally motivate learners to participate, and we explore the effect of presenting the bigger goal in learnersourcing task design.

Research in crowdsourcing has traditionally focused on recruiting paid crowd workers (e.g., Mechanical Turk or oDesk) who are offered monetary reward upon completion of the task. Recent research has begun using voluntary crowds to tackle more domain-specific and

Domain (type of educational video)	Observational / Preliminary Study	User Interface	Data Collection Method	Evaluation
how-to videos - procedural knowledge, structured format	- ToolScape [CHI'13 SRC] - learnersourcing [CHI '13 EA]	- ToolScape [CHI'14] - Crowdy [In preparation for CSCW'15] - Possible large-scale UI work [TODO]	- crowdsourcing [Kim, CHI'14] - active learnersourcing with prompts [In preparation for CSCW'15 with Sarah Weir]	- TODO (Fall 2014)
MOOC videos - conceptual and imperative knowledge, less structured	- data analysis [L@S'14] [L@S '14] [In preparation for TOCHI]	- LectureScape [Kim, in submission] - LectureScape v2 [TODO]	- passive learnersourcing: large-scale data processing [Kim, L@S'14] - active learnersourcing [TODO]	- TODO (Fall 2014)
Office Mix videos - interactive lessons	- TODO (Summer 2014)	- TODO (Summer 2014)	- native plugin support	- TODO (Summer 2014)

Figure 3-1: Current progress summary

complex problems not possible with random crowds. Communitysourcing [19] outsources tasks that community members are qualified and motivated to perform, such as computer science students at a university grading homework problems for free snacks offered from a vending machine. Cobi [29] asks authors of accepted papers at an academic conference to specify papers that are relevant to theirs. The authors' input is used by the system to detect preferences and constraints, which the conference organizers consider when scheduling the conference.

Research on multi-stage crowd workflows inspired the design of our crowdsourcing and learnersourcing methods. Soylent [4] has shown that splitting tasks into the Find-Fix-Verify stages improves the quality and accuracy of crowd workers' results. Other multi-stage crowdsourcing workflows were designed for

Chapter 4

System Designs

This section outlines the systems that will comprise the core of my thesis. I organized the systems around three video types: how-to videos, lecture videos, and interactive lessons. A summary of current progress and plan is described in reffig:progress.

4.1 How-to videos

4.1.1 [Complete] ToolScape: Crowdsourcing Step Information

Millions of learners today use how-to videos to master new skills in a variety of domains. But browsing such videos is often tedious and inefficient because video player interfaces are not optimized for the unique step-by-step structure of such videos. This research aims to improve the learning experience of existing how-to videos with *step-by-step annotations*.

We first performed a formative study to verify that annotations are actually useful to learners. We created ToolScape, an interactive video player that displays step descriptions and intermediate result thumbnails in the video timeline. Learners in our study performed better and gained more self-efficacy using ToolScape versus a traditional video player.

To add the needed step annotations to existing how-to videos at scale, we introduce a novel crowdsourcing workflow. It extracts step-by-step structure from an existing video, including step times, descriptions, and before and after images. We introduce the Find-Verify-Expand design pattern for temporal and visual annotation, which applies clustering,

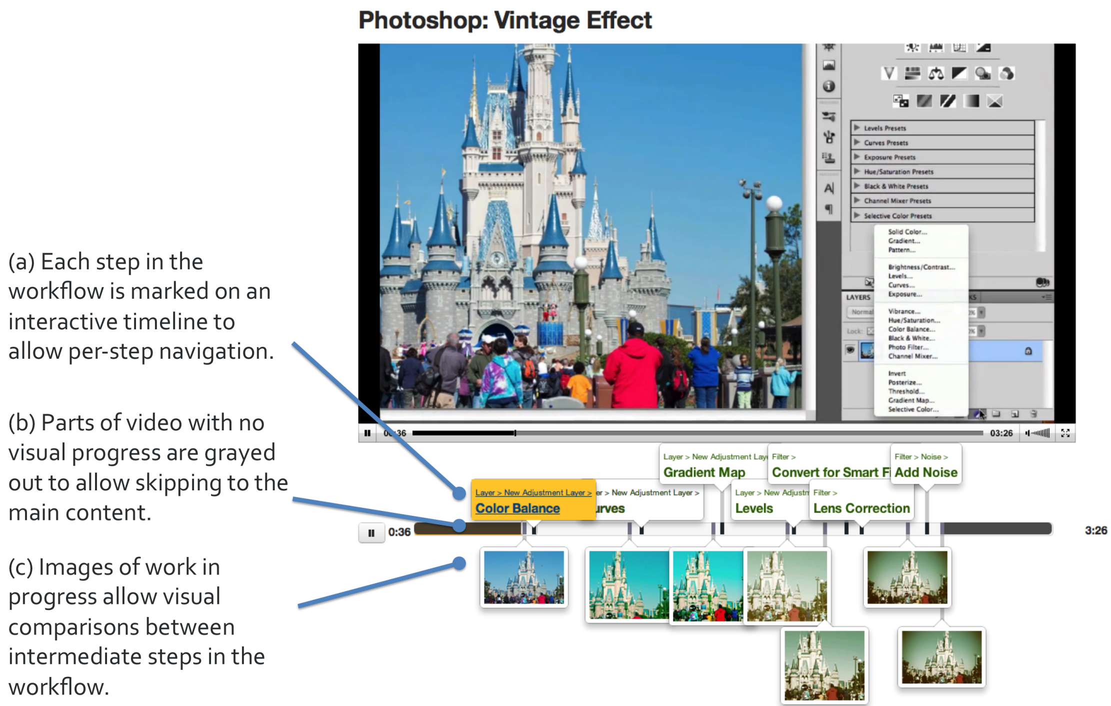

Figure 4-1: ToolScape augments a web-based video player with an interactive timeline. Annotations are shown above the timeline (a), screenshots of intermediate states are shown below the timeline (c), and the gray regions at both ends (b) show "dead times" with no meaningful progress (e.g., waiting for Photoshop to launch).

text processing, and visual analysis algorithms to merge crowd output. The workflow does not rely on domain-specific customization, works on top of existing videos, and recruits untrained crowd workers. We evaluated the workflow with Mechanical Turk, using 75 cooking, makeup, and Photoshop videos on YouTube. Results show that our workflow can extract steps with a quality comparable to that of trained annotators across all three domains with 77% precision and 81% recall.

4.1.2 [Ongoing] Crowdy: Learnersourcing High-Level Goals

We are currently extending the ToolScape workflow in two ways: 1) collect data from motivated learners instead of paid crowd workers, and 2) extract high-level goals from how-to videos instead of low-level steps.

Previous research suggests that users learn more from how-to videos when labels for

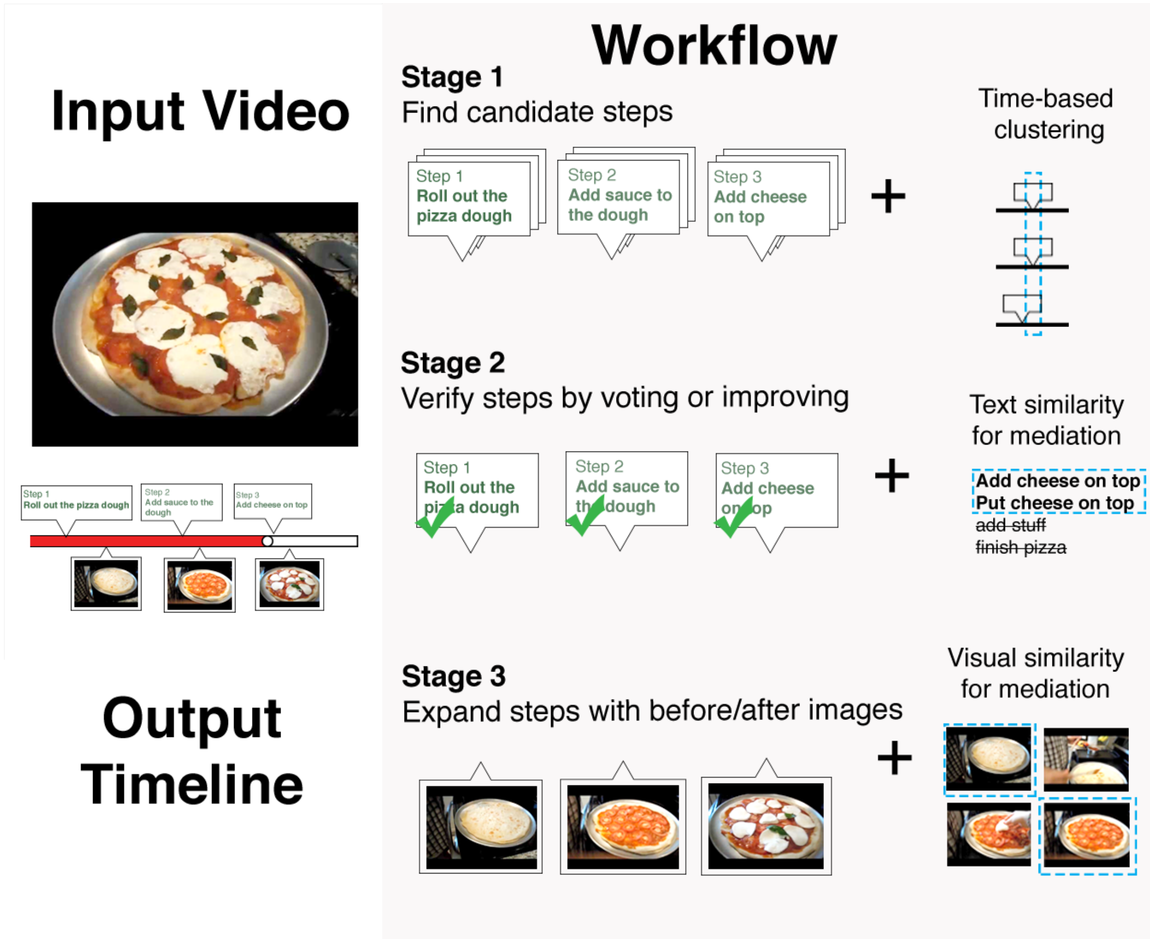

Figure 4-2: Our crowdsourcing workflow extracts step-by-step information from a how-to video with their descriptions and before/after images. It features the Find-Verify-Expand design pattern, time-based clustering, and text/visual analysis techniques. Extracted step information can be used to help learners navigate how-to videos with higher interactivity.

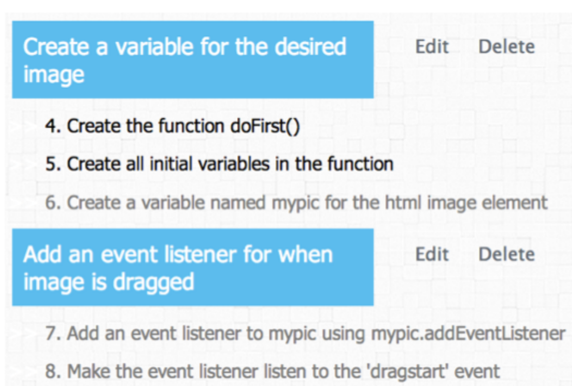

Figure 4-3: Subgoals, shown in blue background, are a cluster of low-level steps that improve learning when presented to learners.

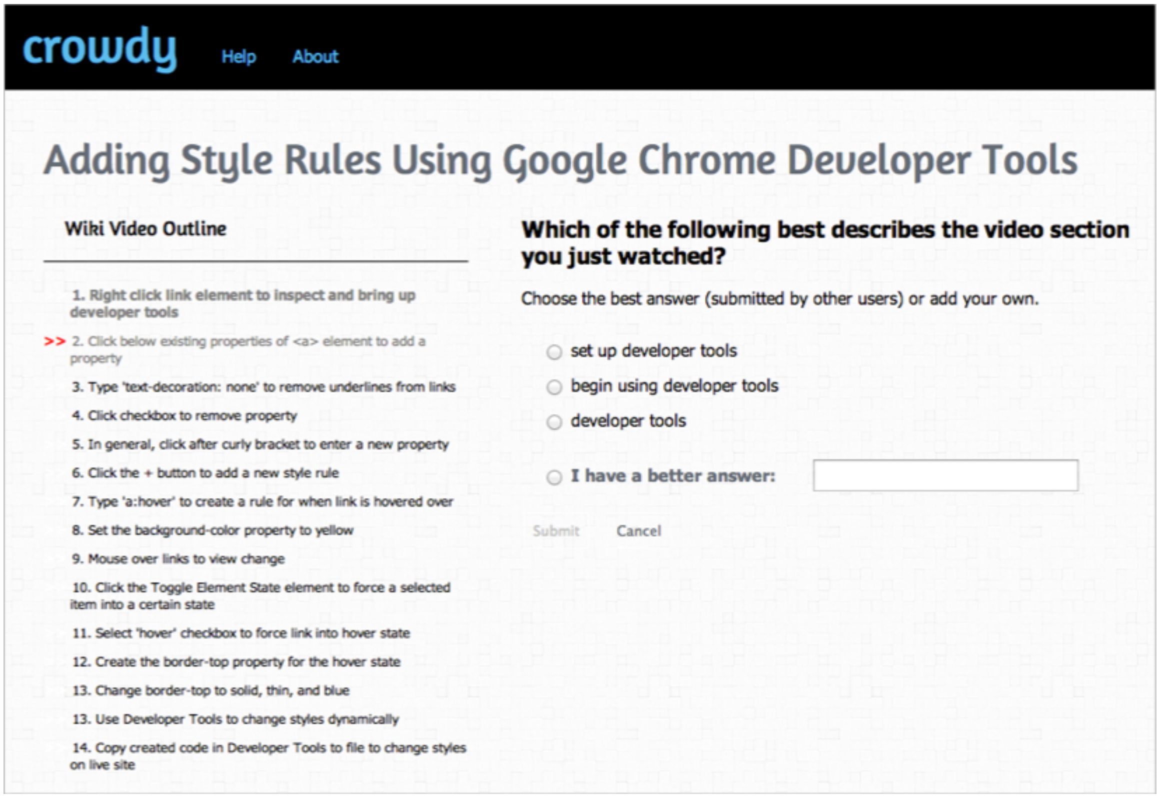

Figure 4-4: An example task in Crowdy, asking the learner to pick the best subgoal for a segment that the learner just watched.

groups of steps (subgoals) are shown [6, 33, 34]. But non-expert, paid crowd workers face difficulty in generating quality subgoals. In order to generate this information, we propose a learnersourcing approach where we gather useful information from people trying to actively learn from a video. To demonstrate this method, we created Crowdy, a workflow and a video player UI that encourages users to contribute and refine subgoals for a given how-to video.

We deployed our video learning interface to students in 6.813, with web programming tutorial videos. Based on the findings from this pilot deployment, we recently released Crowdy publicly. To evaluate Crowdy, we plan to compare the subgoals generated by Crowdy against subgoals generated by experts.

4.1.3 [Proposal] Capstone User Study

We hypothesize that learners will learn better with crowdsourced low-level steps and learnersourced high-level goals. While previous literature has shown the learning benefits of having such information generated by experts or instructors, no research has been done to show similar learning benefits when the information comes from the crowd and learners. We plan to run a laboratory user study specifically designed to measure learning gains when working with labels generated by ToolScape and Crowdy. Furthermore, we hypothesize that the process of generating subgoal labels will have learning benefits as well. The user study will look into the motivational and pedagogical effects of the labeling activity. Running this study will complete a story in the thesis: 1) learners and crowds generate data, 2) a video player UI is automatically built based on the data, and 3) the UI built this way improves learning.

4.1.4 [Optional Proposal] UI for learning with 1000s of annotated videos

Now that we have a scalable method for annotating how-to videos, what can we do if we actually had 1000s of videos fully annotated? We believe we can support more diverse learning patterns by providing step- and goal-level navigation, search, and recommendations, across multiple videos. Imagine that a learner is watching a cooking video where the instructor assumes the learner knows how to poach an egg, but the learner doesn't. What if the UI recommends five snippets from other videos that describe how to poach an egg with varying level of detail? The ToolScape annotation method can be used to mine collective intelligence dispersed in thousands of videos. The role of UI for exploration and navigation is crucial in helping the learner make sense out of the large repository of task examples.

4.2 Lecture Videos

4.2.1 [Ongoing] Exploratory Data Analysis

We started with an exploratory analysis of learners' second-by-second video interaction data from edX MOOCs. With thousands of learners watching the same online lecture

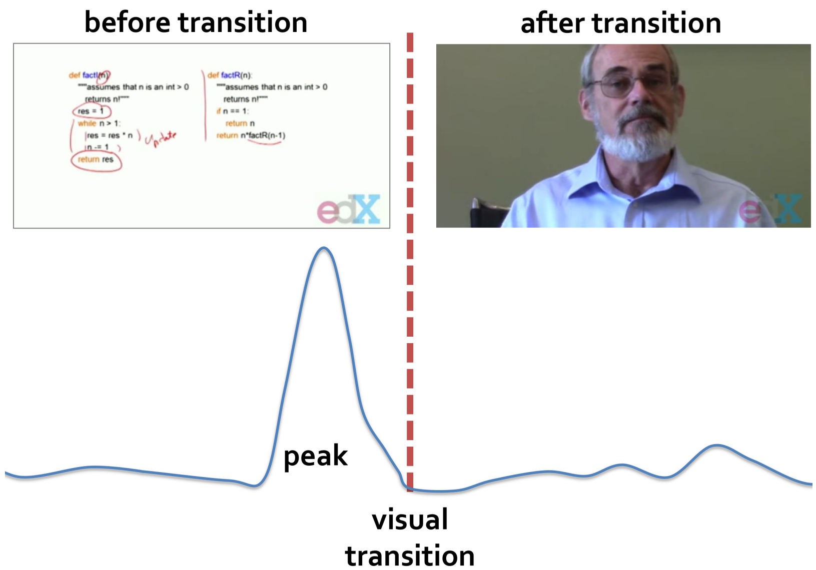

Figure 4-5: An example interaction peak. This peak represents students returning to see the code snippet slide that disappeared after transitioning into the talking head. An abrupt transition might not give students enough time to comprehend what's presented.

videos, analyzing video watching patterns provides a unique opportunity to understand how students learn with videos. We ran a large-scale analysis of in-video dropout and peaks in viewership and student activity, using second-by-second user interaction data from 862 videos in four MOOCs on edX. We found higher dropout rates in longer videos, re-watching sessions (vs first-time), and tutorials (vs lectures). Peaks in re-watching sessions and play events might indicate points of interest and confusion. Results show that tutorials (vs lectures) and re-watching sessions (vs first-time) lead to more frequent and sharper peaks. In attempting to reason why peaks occur by sampling 80 videos, we observed that 61% of the peaks accompany visual transitions in the video, e.g., a slide view to a classroom view. Based on this observation, we identified five student activity patterns that can explain peaks: starting from the beginning of a new material, returning to missed content, following a tutorial step, replaying a brief segment, and repeating a non-visual explanation. This analysis has design implications for video authoring, editing, and interface design, providing a richer understanding of video learning on MOOCs.

We are currently extending this work by analyzing text from transcripts, looking for linguistic patterns and topical transitions around interaction peaks. Combining multiple

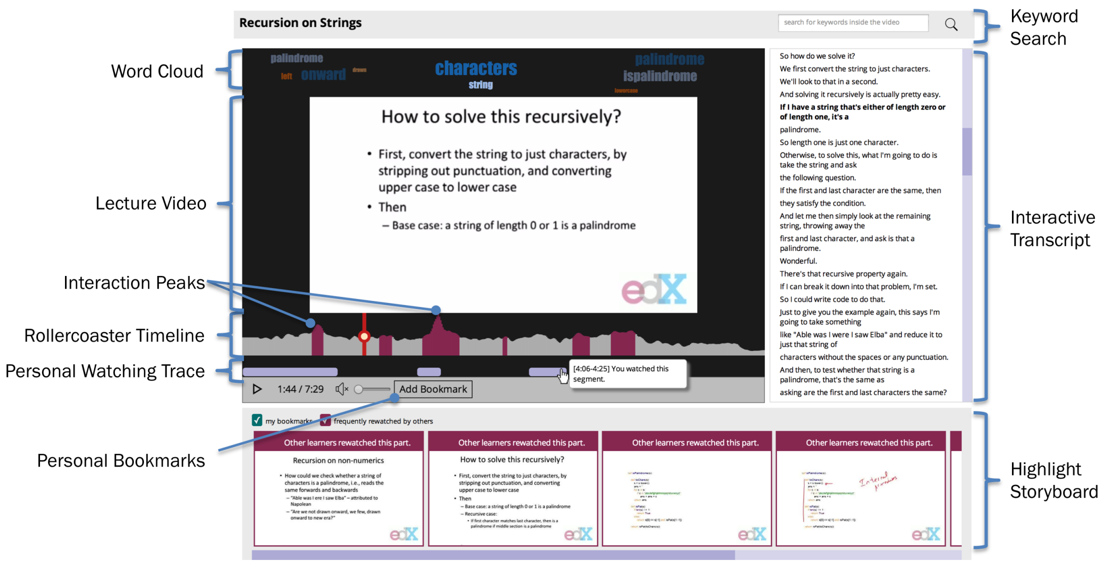

Figure 4-6: LectureScape presents three sets of novel interaction techniques to improve navigation of educational videos. 1) Dynamic timelines (Rollercoaster Timeline, Interaction Peaks, and Personal Watching Trace), 2) Enhanced in-video search (Keyword Search and Interactive Transcript), 3) Highlights (Word Cloud, Personal Bookmarks, Highlight Storyboard). All techniques are powered by interaction data aggregated over all video watchers.

data streams can lead to discovering meaningful video learning patterns otherwise not possible, as each stream brings in a complementary perspective. While our earlier work looked at content type and production style with manually annotated videos, we plan to explore the feasibility of automated methods for interpreting interaction peaks. Such multi-stream, detailed analytics can provide a "debugging" interface for instructors and video editors to improve their videos.

4.2.2 [Complete] LectureScape: Data-Driven Interaction Techniques for Lecture Videos

Interaction data has the potential to help not only instructors to improve their videos, but also to enrich the learning experience of educational video watchers. We explored the design space of data-driven interaction techniques for educational video navigation. We presented a set of techniques that augment existing video interface widgets, including: a 2D video timeline with an embedded visualization of collective navigation traces; dynamic and non-linear timeline scrubbing; data-enhanced transcript search and keyword summary; automatic display of relevant still frames next to the video; and a visual summary representing points with high learner activity. To evaluate the feasibility of the techniques, we ran a laboratory user study with simulated learning tasks. Learners rated watching lecture videos with interaction data to be efficient and useful in completing the tasks, enabling them to employ more diverse navigation patterns. Generally, learners' comments about LectureScape were positive. Learners noted that "It's not like cold-watching. It feels like watching with other students.", and "[interaction data] makes it seem more classroom-y, as in you can compare yourself to what how other students are learning and what they need to repeat."

4.2.3 [Proposal] Live Deployment of LectureScape in a MOOC

To evaluate the passive learnersourcing method used in enabling LectureScape, we plan to deploy LectureScape in a live MOOC. There are research questions around the role of collective data when no existing data is available: When do interaction patterns start to emerge? Would small trends in the beginning skew further learners' interaction patterns? Furthermore, we hypothesize that keeping track of both collective and personal interaction traces can lead to better learning, when compared against having only collective or personal traces alone. We plan to test this hypothesis through a laboratory study.

4.3 Interactive Lessons

4.3.1 [Proposal] New Formats for Instructional Videos

Both the how-to and lecture video research threads have focused on reverse-engineering existing videos and interaction data to discover useful structure for navigation and learning. A missing piece is, what if we had access to the video authoring tool? Can we encourage instructors to create better materials in the first place, and can we design richer learnersourcing prompts? Furthermore, can we radically rethink how instructional videos are structured, delivered, and presented?

During my internship at Microsoft Research in the summer of 2014, I plan to answer these questions with Office Mix, a Powerpoint plugin for instructors to create interactive lessons. It allows instructors to add talking heads, voice over, and free-form drawing to a slide, and also to embed interactive quizzes and Khan Academy videos. Using this platform, I plan to experiment various interventions to foster higher quality content generation and deeply integrated learnersourcing prompts.

Chapter 5

Status and Timeline

The following are published or submitted papers that are marked as "complete" in the previous section.

- ToolScape: CHI 2013 EA [24], CHI 2014 [28]
- Crowdy: CHI 2013 EA [27], CSCW 2015 (to be submitted in June 2014, fallback: CHI 2015)
- Exploratory data analysis: Learning at Scale 2014 [17, 25], CHI 2014 workshop [26], TOCHI (to be submitted in May 2014)
- LectureScape: UIST 2014 (in review, fallback: CHI 2015)

The following is a timeline until graduation with major milestones.

5.1 Timeline: Graduate in June 2015

- Summer 2014: Internship at Microsoft Research
- Sep 2014: Submit to CHI 2015 - summer work at MSR
- Sep 2014: Submit to CHI 2015 - the how-to video capstone user study
- Oct 2014: Submit to Learning at Scale 2015 - the LectureScape deployment

- Winter 2014: Faculty application
- Spring 2015: Faculty interviews
- May 2015: Thesis defense
- June 2015: Graduation

Chapter 6

Contributions

This thesis will demonstrate that active and passive learnersourcing can be used to improve video learning. We believe that learnersourcing can provide navigational, social, and pedagogical benefits to learners who are watching the video, while generating useful information for future learners as a byproduct. We will make contributions in three areas: 1) extracting hierarchical solution structure from how-to videos, 2) leveraging collective interaction traces to create data-driven interaction techniques for lecture videos, and 3) novel instructional video formats that natively support learnersourcing.

Bibliography

[1] Gregory D. Abowd. Classroom 2000: An experiment with the instrumentation of a living educational environment. *IBM systems journal*, 38(4):508–530, 1999.

[2] Abir Al-Hajri, Gregor Miller, Matthew Fong, and Sidney S. Fels. Visualization of personal history for video navigation. In *CHI '14*, 2014.

[3] F. Arman, R. Depommier, A. Hsu, and M.-Y. Chiu. Content-based browsing of video sequences. In *MULTIMEDIA '94*, pages 97–103, 1994.

[4] Michael S. Bernstein, Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. Soylent: a word processor with a crowd inside. In *UIST '10*, pages 313–322, 2010.

[5] Axel Carlier, Vincent Charvillat, Wei Tsang Ooi, Romulus Grigoras, and Geraldine Morin. Crowdsourced automatic zoom and scroll for video retargeting. In *Multimedia '10*, pages 201–210, 2010.

[6] Richard Catrambone. The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, 127(4):355, 1998.

[7] Michelene TH Chi and R Wylie. Icap: A hypothesis of differentiated learning effectiveness for four modes of engagement activities. *Educational Psychologist*, to appear.

[8] Pei-Yu Chi, Sally Ahn, Amanda Ren, Mira Dontcheva, Wilmot Li, and Björn Hartmann. Mixt: Automatic generation of step-by-step mixed media tutorial. In *UIST '12*, 2012.

[9] Pei-Yu (Peggy) Chi, Joyce Liu, Jason Linder, Mira Dontcheva, Wilmot Li, and Björn Hartmann. Democut: generating concise instructional videos for physical demonstrations. In *UIST '13*. ACM, 2013.

[10] Konstantinos Chorianopoulos. Collective intelligence within web video. *Human-centric Computing and Information Sciences*, 3(1):10, 2013.

[11] Scotty D Craig, Michelene TH Chi, and Kurt VanLehn. Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of Educational Psychology*, 101(4):779, 2009.

[12] Erika L Ferguson and Mary Hegarty. Learning with real machines or diagrams: application of knowledge to real-world problems. *Cognition and Instruction*, 13(1):129–160, 1995.

[13] Jennifer Fernquist, Tovi Grossman, and George Fitzmaurice. Sketch-sketch revolution: an engaging tutorial system for guided sketching and application learning. In *UIST '11*, pages 373–382, 2011.

[14] A. Girgensohn and J. Boreczky. Time-constrained keyframe selection technique. In *Multimedia Computing and Systems*, volume 1, pages 756–761 vol.1, 1999.

[15] Floraine Grabler, Maneesh Agrawala, Wilmot Li, Mira Dontcheva, and Takeo Igarashi. Generating photo manipulation tutorials by demonstration. In *SIGGRAPH '09*, pages 1–9, 2009.

[16] Tovi Grossman, Justin Matejka, and George Fitzmaurice. Chronicle: capture, exploration, and playback of document workflow histories. In *UIST '10*, 2010.

[17] Philip J. Guo, Juho Kim, and Rob Rubin. How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 41–50, New York, NY, USA, 2014. ACM.

[18] Rassule Hadidi and Chung-Hsien Sung. Students' acceptance of web-based course offerings: an empirical assessment. *AMCIS 1998*, 1998.

[19] Kurtis Heimerl, Brian Gawalt, Kuang Chen, Tapan Parikh, and Björn Hartmann. Communitysourcing: Engaging local crowds to perform expert work via physical kiosks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1539–1548, New York, NY, USA, 2012. ACM.

[20] William C. Hill, James D. Hollan, Dave Wroblewski, and Tim McCandless. Edit wear and read wear. In *CHI '92*, pages 3–9, 1992.

[21] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In *CVPR '07*, pages 1–8, 2007.

[22] Amy Hurst, Jennifer Mankoff, Anind K. Dey, and Scott E. Hudson. Dirty desktops: Using a patina of magnetic mouse dust to make common interactor targets easier to select. In *UIST '07*, pages 183–186, 2007.

[23] Dan Jackson, James Nicholson, Gerrit Stoeckigt, Rebecca Wrobel, Anja Thieme, and Patrick Olivier. Panopticon: A parallel video overview system. In *UIST '13*, pages 123–130, 2013.

[24] Juho Kim. Toolscape: Enhancing the learning experience of how-to videos. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 2707–2712, New York, NY, USA, 2013. ACM.

[25] Juho Kim, Philip J. Guo, Daniel T. Seaton, Piotr Mitros, Krzysztof Z. Gajos, and Robert C. Miller. Understanding in-video dropouts and interaction peaks inonline lecture videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 31–40, New York, NY, USA, 2014. ACM.

[26] Juho Kim, Shang-Wen (Daniel) Li, Carrie J. Cai, Krzysztof Z. Gajos, and Robert C. Miller. Leveraging video interaction data and content analysis to improve video learning. In *CHI '14 Workshop on Learning Innovations at Scale*, 2014.

[27] Juho Kim, Robert C. Miller, and Krzysztof Z. Gajos. Learnersourcing subgoal labeling to support learning from how-to videos. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 685–690, New York, NY, USA, 2013. ACM.

[28] Juho Kim, Phu Tran Nguyen, Sarah Weir, Philip J. Guo, Robert C. Miller, and Krzysztof Z. Gajos. Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 4017–4026, New York, NY, USA, 2014. ACM.

[29] Juho Kim, Haoqi Zhang, Paul André, Lydia B. Chilton, Wendy Mackay, Michel Beaudouin-Lafon, Robert C. Miller, and Steven P. Dow. Cobi: A community-informed conference scheduling tool. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 173–182, New York, NY, USA, 2013. ACM.

[30] Benjamin Lafreniere, Tovi Grossman, and George Fitzmaurice. Community enhanced tutorials: improving tutorials with multiple demonstrations. In *CHI '13*, pages 1779–1788, 2013.

[31] Yong Jae Lee, C. Lawrence Zitnick, and Michael F. Cohen. Shadowdraw: real-time user guidance for freehand drawing. *ACM Trans. Graph.*, 30(4):27:1–27:10, July 2011.

[32] Francis C. Li, Anoop Gupta, Elizabeth Sanocki, Li-wei He, and Yong Rui. Browsing digital video. In *CHI '00*, pages 169–176, 2000.

[33] Lauren E Margulieux, Richard Catrambone, and Mark Guzdial. Subgoal labeled worked examples improve k-12 teacher performance in computer programming training. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Austin, TX, USA, 2013.

[34] Lauren E. Margulieux, Mark Guzdial, and Richard Catrambone. Subgoal-labeled instructional material improves performance and transfer in learning to develop mobile applications. In *Proceedings of the Ninth Annual International Conference on International Computing Education Research*, ICER '12, pages 71–78, New York, NY, USA, 2012. ACM.

[35] Justin Matejka, Tovi Grossman, and George Fitzmaurice. Patina: Dynamic heatmaps for visualizing application usage. In *CHI '13*, pages 3227–3236, 2013.

[36] Justin Matejka, Tovi Grossman, and George Fitzmaurice. Swifter: Improved online video scrubbing. In *CHI '13*, pages 1159–1168, 2013.

[37] Richard E Mayer. Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, pages 31–48, 2005.

[38] Richard E Mayer and Roxana Moreno. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1):43–52, 2003.

[39] Robert Mertens, Rosta Farzan, and Peter Brusilovsky. Social navigation in web lectures. In *HYPERTEXT '06*, pages 41–44, 2006.

[40] Mathieu Nancel and Andy Cockburn. Causality: A conceptual model of interaction history. In *CHI '14*, 2014.

[41] James Nicholson, Mark Huber, Daniel Jackson, and Patrick Olivier. Panopticon as an elearning support search tool. In *CHI '14*, 2014.

[42] Dan R Olsen and Brandon Moon. Video summarization based on user interaction. In *EuroITV '11*, pages 115–122, 2011.

[43] Suporn Pongnumkul, Mira Dontcheva, Wilmot Li, Jue Wang, Lubomir Bourdev, Shai Avidan, and Michael Cohen. Pause-and-play: Automatically linking screencast video tutorials with applications. In *UIST 2011*, 2011.

[44] E.F. Risko, T. Foulsham, S. Dawson, and A. Kingstone. The collaborative lecture annotation system (clas): A new tool for distributed learning. *Learning Technologies, IEEE Transactions on*, 6(1):4–13, 2013.

[45] Ryan Shaw and Marc Davis. Toward emergent representations for video. In *Multimedia '05*, pages 431–434, 2005.

[46] Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. Animation: can it facilitate? *International journal of human-computer studies*, 57(4):247–262, 2002.

[47] Luis Von Ahn. Games with a purpose. *Computer*, 39(6):92–94, 2006.

[48] Luis von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 319–326, New York, NY, USA, 2004. ACM.

[49] Jude Yew, David A. Shamma, and Elizabeth F. Churchill. Knowing funny: genre perception and categorization in social video sharing. In *CHI '11*, pages 297–306, 2011.

[50] Dongsong Zhang, Lina Zhou, Robert O. Briggs, and Jay F. Nunamaker Jr. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management*, 43(1):15 – 27, 2006.