Delta: A Tool For Representing and Comparing Workflows

Nicholas Kong[*†]**, Tovi Grossman**[*]**, Björn Hartmann**[†]**, George Fitzmaurice**[*]**, Maneesh Agrawala**[†]

[*]Autodesk Research, Toronto, Ontario, Canada
{*tovi.grossman,george.fitzmaurice*}@autodesk.com

[†]University of California at Berkeley, USA
{*nkong,bjoern,maneesh*}@eecs.berkeley.edu

ABSTRACT

Tutorials and sample workflows for complicated, feature-rich software packages are widely available online. As a result users must differentiate between workflows to choose the most suitable one for their task. We present *Delta*, an interactive workflow visualization and comparison tool that helps users identify the tradeoffs between workflows. We conducted an initial study to identify the set of attributes users attend to when comparing workflows, finding that they consider result quality, their knowledge of commands, and the efficiency of the workflow. We then designed Delta to surface these attributes at three granularities: a high-level, clustered view; an intermediate-level list view that contains workflow summaries; and a low-level detail view that allows users to compare two individual workflows. Finally, we conducted an evaluation of Delta on a small corpus of 30 workflows and found that the intermediate list view provided the best information density. We conclude with thoughts on how such a workflow comparison system could be scaled up to larger corpora in the future.

Author Keywords

Workflow comparison; Tutorials; Search interfaces; Delta.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

INTRODUCTION

Design software, such as image editors, animation tools, and 3D modeling applications, provide users with powerful ways to create and modify virtual content. Despite advances in HCI research, the interfaces for such applications are complex and can be difficult to learn [12]. While software documentation can provide assistance with locating and using individual components, users often refer to online tutorials to learn the required workflows for completing higher-level tasks. For example, a Photoshop user wishing to correct red-eye in a photograph could search online to find step-by step instructions.

Today, there are an abundance of online communities and websites specifically dedicated to publishing such *workflows* – sequences of commands required to

accomplish a defined task. Furthermore, researchers have begun to build systems that automatically record user action streams, which can be used by others to learn a workflow [3, 8, 10, 11]. As the corpora of these online learning resources continue to grow, users are likely to find multiple workflows for the same task. Workflows can vary in the tools they use, the quality of the results, or may have differing prerequisites (such as requiring specific plug-ins).

The challenge for a user is no longer searching for and finding *any* workflow – the challenge is identifying the *most suitable* workflow from potentially hundreds of possible candidates. In addition, individual websites tend to use a wide range of layouts and formats for presenting workflows, potentially making it difficult for users to identify differences in the workflows themselves. To date, there is little research investigating how users explore and identify possible workflows to accomplish a task when multiple candidates exist. Furthermore, no tools have been explicitly designed to support the process of exploring and comparing workflows.

In this paper we focus on image-editing workflows for Photoshop, because it is a well-known design application with a large number of tutorials available online. We first perform a study to understand users' workflow searching behaviors, and identify the criteria they use when choosing between multiple image-editing workflows. Our main finding is that users primarily consider the quality of the result, their knowledge of commands (and thus their effects), and the efficiency of the workflow.

The results of this study guide us in the design of *Delta*, a new interactive tool to aid the process of workflow comparison and exploration. Given a corpus of up to 50 workflows that describe the same task, our tool allows users to understand the scope of available workflows; explore each of the workflows in detail; and understand the similarities and differences in the structure (i.e., command sequence) of individual workflows.

Our tool allows users to visualize and compare workflows at three levels of granularity: a high-level clustered view; an intermediate-level list view that contains workflow summaries; and a low-level detail view that allows users to compare two individual workflows. To ease comparisons, the tool displays all workflows with a uniform presentation format within each of the three granularity levels.

To understand the potential benefits of the tool, and to evaluate the different comparison granularities which it provides, we conducted a user study. In comparison to a

traditional web search interface, users preferred our tool, and in particular its ability to easily compare workflows. Observations and comments also indicated that the intermediate list view provided the ideal information density for most of the workflow comparison tasks. We conclude by describing limitations, design refinements, and implications for workflow browsing.

RELATED WORK
Our system draws on related work from three main areas: models for how users find information, search/exploration interfaces, and workflow recording systems.

Information Scent and Finding Information
Delta is fundamentally an information search tool. Some researchers have developed theoretical models for how people find information. Information foraging [22] has been applied to model and predict user behavior in web search [4, 5]. A key concept of information foraging theory is information scent, or the contextual cues that allow users to assess the benefits and costs of distal (i.e., hidden) content [4]. In this work, we identify the contextual cues that users find most effective when they search for workflows.

Search Interfaces for Large Corpora
Workflows are a specific type of structured data. Researchers have developed systems for searching or accessing large collections of unstructured data, such as documents [14], images [16], or web pages for design [20]. Others have looked at systems for exploring more structured data. One example is VisComplete, a dataflow suggestion system built within a visual programming system for visualization [17]. As users create a dataflow out of a sequence of function modules, the system makes suggestions of other dataflows that are similar to the current dataflow. Our tool is designed to accommodate workflows, and also uses similarity metrics to suggest similar workflows in the context of exploring the entire corpus.

Closest to our work is AdaptableGIMP [19], which integrates image-editing workflow search into GIMP. Workflows are stored on a wiki and users can access these wiki pages from within GIMP. They can also directly invoke GIMP commands from the wiki pages. In contrast, our tool is designed to help people compare workflows, in addition to aiding search tasks.

Workflow Recording Systems
Finally, many researchers have created systems for recording workflows. These systems are complementary to our tool, as their output can be fed into our system. Much of the work in this area focused on recording interaction histories in domains such as painting applications [18], visual analysis [15], and GUI applications [21]. More recent work has focused on enabling users to effectively explore longer workflows by chunking operations. Chronicle [11] and MeshFlow [8] allow users to replay and explore workflows for an image manipulation application and a mesh creation system, respectively. The proliferation of such systems enhances the need for tools to explore workflow corpora.

TERMINOLOGY
In this section, we formally define terms that we use throughout the paper.

A **command** is a user-invoked operation in an application. Examples of Photoshop commands include the activation of palette tools, such as the brush tool, or menu options that open dialog boxes, such as the image level tool. The complexity of a command can vary from inverting the image to manually specifying a color curve.

A **workflow** is a sequence of commands and can take multiple representations. One common representation is a **tutorial**, which describes a workflow in a sequence of steps, frequently with image or video illustrations and explanations for why certain commands are used. Workflows can vary greatly in their length. Our tool has been designed to work with relatively short workflows – anywhere from 5 to 30 steps long – as this is the range of steps we observed while sampling a collection of online tutorials for common image-editing tasks.

STUDY 1: HOW USERS COMPARE WORKFLOWS
To better understand which workflow attributes users consider most important when comparing and assessing workflows, we ran a small study. We used the results of this study in the design of Delta to emphasize the most important attributes. Users looked at a set of workflows that described the same task and chose the one that they would most likely use.

Workflow Representations
We used three different workflow presentation styles. While our goal was not to compare the presentation styles themselves, we reasoned that each presentation style may prompt users to consider different attributes when comparing workflows. Figure 1 shows the three presentation styles we used: *text, text+image*, and *graph*. The *text* representation is a paragraph of text with inline tool icons. It deemphasizes the length of the workflow and the quality of the result while surfacing palette tools using icons. By removing clearly delineated steps and images, it also makes it more difficult for users to skim the workflow, forcing them to read closely.

The *text+image* representation is a step-by-step description annotated with images illustrating each step. It also contains before and after images, which allows users to assess the result quality, and it emphasizes the length of the workflow by numbering steps. Its layout closely followed the tutorials generated in Grabler et al.'s tutorial recording system [10].

Finally, the *graph* representation is a visualization that displays commands as nodes in a graph. It emphasizes the commands and parameters used in the workflow. Each node displays canvas state, similar to Chen et al.'s DAG representation of workflows [3].
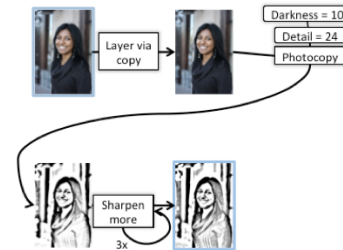
Workflow 1
Create a new layer using Layer via Copy (▣). This new layer will be called *Layer 1*. Apply the Photocopy filter. Set Detail to 24 and Darkness to 10. Finally, apply the Sharpen More filter three times.

(a) Text
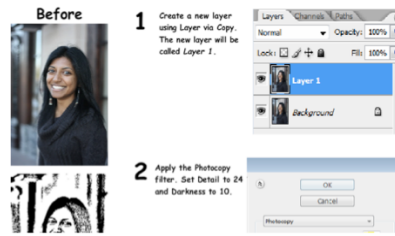
(b) Text+image

(c) Graph

Figure 1. Workflow representations in the first study. (a) The *text* condition shows a workflow as a paragraph with inline tool icons (if they exist). The icon here is added for illustration. (b) The *text+image* condition displays before and after images and individual steps. (c) The *graph* condition displays the canvas after each command. Command names are shown as regular rectangles, parameters as rounded rectangles.

Study Design and Procedure

The study consisted of three trials, one for each representation. In each trial, we showed participants three different workflows for the same Photoshop task. We gave participants a target image and asked them to think aloud while selecting one workflow they would choose to apply to the target image.

In each trial, participants saw a different task: red eye correction, applying a sketch effect, and image sharpening. We chose tasks that were short but complex enough to permit multiple different approaches. We obtained tutorials from the web and transcribed workflows manually. Each participant saw every representation, and no two participants saw the same representation/task pair. We recorded audio during each of the trials.

After the participants had completed all three tasks, they completed a short questionnaire where they scored a list of workflow attributes (on a 5 point Likert scale) on how important the attributes were when comparing or choosing between workflows. The list of attributes, shown in Figure 2, was based on our experience. Although our focus was on the workflow attributes users chose, we also had users rank the effectiveness of the representations based on how clearly they displayed the attributes.

Participants

We recruited six participants through an online classified posting. We had four male and two female participants, who ranged from 19 to 40 years old and had between three and seven years of Photoshop experience.

Results and Discussion

Although we only had six participants, we encountered a diverse cross-section of Photoshop users, whose aims varied when choosing a workflow. For example, P2 was very interested in learning new Photoshop features and would occasionally browse for tutorials to improve his knowledge. In contrast, P3 was most interested in finding a quick and efficient way to complete the task. However, we identified overlap in the workflow attributes participants used when selecting a workflow.

Five of six participants found the *text+image* representation most useful at surfacing almost all of the attributes; one participant preferred the *text* and *graph* representations, saying that the *text+image* representation "*seemed like too much to have to look through.*"

Preferred Workflow Attributes

Figure 2 shows user-rated importance of workflow attributes when comparing workflows. None of the differences in the means are significant after a Bonferroni corrected paired t-test However, users' comments during the think-aloud portion of the study corroborated the general trend of these rankings.

During the think-aloud trials, all of the participants mentioned *quality of the result*, *knowledge of the commands*, and *workflow efficiency* (e.g., length). The second attribute category subsumes a number of workflow attributes: when speaking about commands, users referenced their *familiarity* with the commands and their knowledge of the commands' effects, which is an indicator of a workflow's *flexibility*. Although users rated "the workflow can be used in other situations" as an important attribute, no user mentioned this in the think-aloud. Users may have interpreted this attribute to mean that the workflow is flexible, or that it can apply widely to a variety of images (e.g., a sharpening workflow that would work equally well on a landscape photo and a portrait).

Perhaps unsurprisingly, users considered holistic attributes of workflows (e.g., the quality of the result or the efficiency of the workflow) more important than mechanistic attributes (e.g., number of mouse-clicks or number of dialog boxes). No users mentioned either of these mechanistic attributes during the think-aloud.

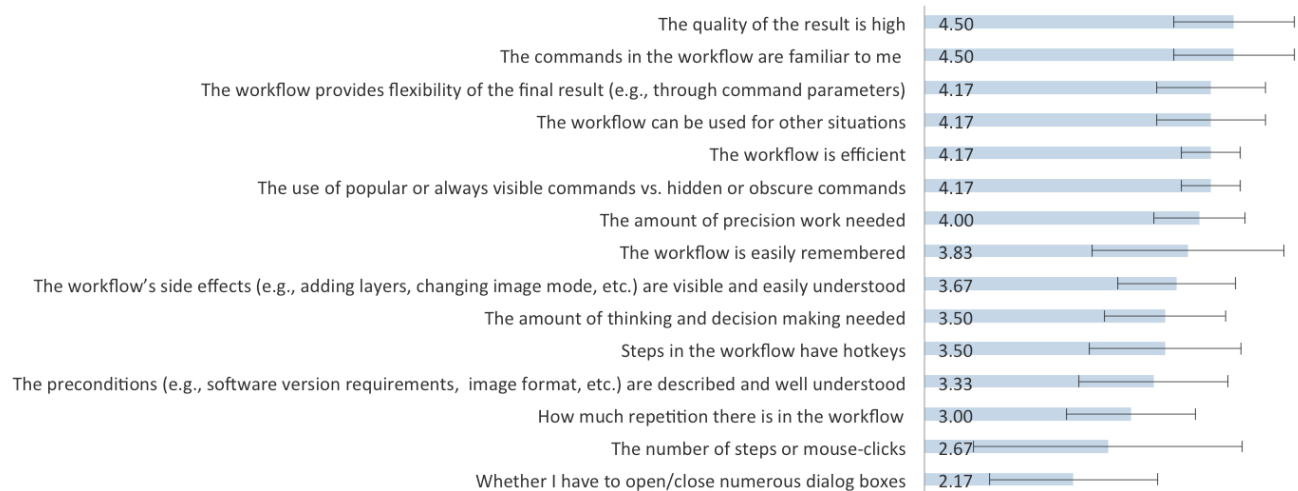|	
The quality of the result is high	4.50
The commands in the workflow are familiar to me	4.50
The workflow provides flexibility of the final result (e.g., through command parameters)	4.17
The workflow can be used for other situations	4.17
The workflow is efficient	4.17
The use of popular or always visible commands vs. hidden or obscure commands	4.17
The amount of precision work needed	4.00
The workflow is easily remembered	3.83
The workflow's side effects (e.g., adding layers, changing image mode, etc.) are visible and easily understood	3.67
The amount of thinking and decision making needed	3.50
Steps in the workflow have hotkeys	3.50
The preconditions (e.g., software version requirements, image format, etc.) are described and well understood	3.33
How much repetition there is in the workflow	3.00
The number of steps or mouse-clicks	2.67
Whether I have to open/close numerous dialog boxes	2.17

Figure 2. Likert-scale rankings for the importance of workflow attributes when comparing workflows. Users rated attributes on a scale from 1-5, where 1 was "not important at all" and 5 was "very important". Bars indicate 95% confidence intervals.

The attributes related to the quality of the result and knowledge of the tools are not independent: users were better able to anticipate the result of a workflow if they were familiar with the tools and their effects, and how they could be parameterized to change the result. This was particularly apparent in the text representation, since it did not contain any images. For example, in the image sharpening task shown in a text representation, P2 chose workflow 1, explaining that:

"... with workflow 1 ... when you're playing around with your 'find edges' and 'levels' and everything, it's only going to pick up what you're looking for, which would probably just be the squirrel in the image."

In this case, P2 used his prior knowledge of the effects of the commands and how they would affect the target image. Likewise, P5 noted that the before/after images for the workflows in the sharpening task were too small to see differences, and so chose a workflow using his "experience", saying:

"...[the workflow] allows you to have more control over the important edges. Gives you better contrast control."

Finally, all users mentioned workflow length in the think-aloud, which is related to the efficiency of the workflow. However, only one participant, P3, placed a larger emphasis on the length of the workflow rather than the result quality. For example, in the sketch effect task displayed as text, P3 selected workflow 1 as it was "*pretty straightforward, pretty efficient in terms of time*," even though she noted that workflow 2 might produce a better result but was too "*labor-intensive*."

Other users thought efficiency was important, but took efficiency to mean that the workflow produces a good result and does not contain any extraneous commands, even if the commands in the workflow require many clicks to execute (e.g., specifying a color curve). An efficient workflow may contain many commands as long as each is necessary to achieve the final result. For example, P6 chose a longer sharpening workflow because of *"...familiarity. ... though there are a lot of steps involved, it provides the most specification from my perspective."*

Tutorial Seeking Behaviors
We were also interested in users' current practice to find workflows. We asked participants how often they used tutorials and where they typically found them. All of our participants used search engines to find tutorials. One also used books, and another also used magazines. However, they both reported that their predominant method to seek out tutorials was web search. One of our participants preferred video tutorials, while another disliked them; the others did not have a preference for the modality of the tutorial. Finally, users reported that they selected tutorials based on sites originating from a trusted domain (e.g., adobe.com), and search engine ranking. These results corroborate prior work on information seeking behavior via search engines [6, 9, 13].

Implications for System Design
The results of this study provide some guidance on the design of a tool to support workflow comparison, and in particular how workflows should be represented. A consistent and compact representation of a workflow that displays the most important information (i.e., provides the strongest information scent [22]) will make it easy for users to compare collections of workflows, or two specific workflows. The main implications from the initial study are as follows:

• *Images of the results and intermediate steps are important.* Users wanted to not only see the before/after images for each workflow, but also images illustrating in-between steps to help them visualize operations.

- *Commands are important.* Users used their knowledge of commands and their effects to assess how effective or flexible a workflow was (i.e., how the parameters of commands could be changed to affect the result).

- *Workflow efficiency is important.* Some users wanted to know how quick or efficient a workflow was. A workflow comparison tool should display the number of commands or steps as a proxy for this information.

METRICS FOR COMPARING WORKFLOWS

Ideally, a workflow comparison tool would be able to algorithmically suggest similar or different workflows to the user. Some systems designed for the exploration of large corpora of unstructured data, such as documents or images, do this by using clustering ([7, 16]). One challenge for creating a clustering interface for workflows is quantifying how similar two workflows are; once this metric is defined, we can use it to create clusters of similar workflows. However, workflows could be grouped using different notions of similarity: for example, workflows could be grouped by result quality (e.g., [20]), or by how many commands have been used by the user. Similarity metrics can also be used to rank workflows. In this section, we explore the space of similarity metrics and how they relate to the attributes of workflows users deemed important in the first study.

In our first study, we identified three workflow attribute types users consider when comparing workflows: *quality of the result*, *knowledge of the tools*, and *workflow efficiency*. For each of these attribute types, we present potential similarity metrics that could be used to cluster, or rank, a workflow collection. A similarity metric could also be a combination of metrics from these types.

Result Similarity

Workflows may be grouped by the results they produce, which would help users choose the workflow with the desired quality. Similarity metrics of this type are more difficult to operationalize than metrics of other types, as result quality is a subjective judgment. However, result similarity metrics may be amenable to crowdsourcing techniques [24]. We could also use metrics such as PageRank or popularity ratings such as social media ratings (Facebook likes, Google+ +1s, etc.); PageRank is already widely used as a ranking metric given web search's popularity for finding tutorials.

A user might also wish to find a workflow that matches their task. For example, a user may wish to remove red eye from a photo of a child, and so would want to find a red eye removal workflow that also uses a photo of a child. A result similarity metric based on image attributes, such as the objects in the photo (e.g., faces, cars, etc.) or lower-level attributes such as color, would be useful in this scenario.

Command Similarity

Another way to quantify workflow similarity is by using the commands or command sequences used. This class of metrics would help surface what workflows contained familiar commands, or which allowed the user to flexibly modify the final result. Potential metrics include the number of shared commands; a workflow familiarity rating based on a users' personal usage profile; or the edit distance [25] between two sequences of commands. For our system, we used an edit-distance metric, described later.

Workflow Efficiency

Although it was a lesser concern for participants, the efficiency of a workflow was identified as an important differentiating attribute. Explicitly determining the efficiency of a workflow would be challenging. However, grouping or ranking workflows by their length, either in the time taken or the number of commands used, may serve as a suitable approximation. For our system, we use the number of commands to rank workflows.

DELTA: APPLICATION DESIGN

In this section, we describe the design of our new workflow exploration tool, *Delta*. Based on study participants' reports and our own experience, we designed Delta's interface to explore corpora of that are representative of collections of image manipulation tutorials online: Delta targets collections of several dozen workflows, each containing up to 30 commands. Three complementary views provide information about the workflow corpus at different levels of detail (Figure 3): a *cluster view* presents a high-level overview of the entire corpus; a *list view* provides workflow summaries; a *detail view* lists individual commands and enables pairwise comparisons. We describe each in turn.

Cluster View

The cluster view provides a high-level summary of the corpus by grouping workflows by similarity. Clusters enable users to rapidly distinguish between similar and different workflows.

We selected a command similarity metric for clustering – it can be computed without requiring additional metadata such as ratings, and it is closely related to the *knowledge of commands* workflow attribute. We computed similarity by modeling each workflow as a sequence of commands, then computing the *edit distance* [25] between sequences of commands. The edit distance is the sum of the costs of the insertions, deletions, and substitutions that are required to transform one command sequence into another. We used the distance between commands in the Photoshop menu hierarchy to set the substitution cost: for example, substituting Gaussian Blur for Motion Blur incurs a low substitution cost, because both are siblings in the Blur menu. We used hierarchical agglomerative clustering [23] to create clusters, although other methods (such as k-means) are possible. Finally, we laid out the clusters using the quantum treemap algorithm [2], as it was designed to compactly lay out items with a minimum display dimension, such as our workflow images.
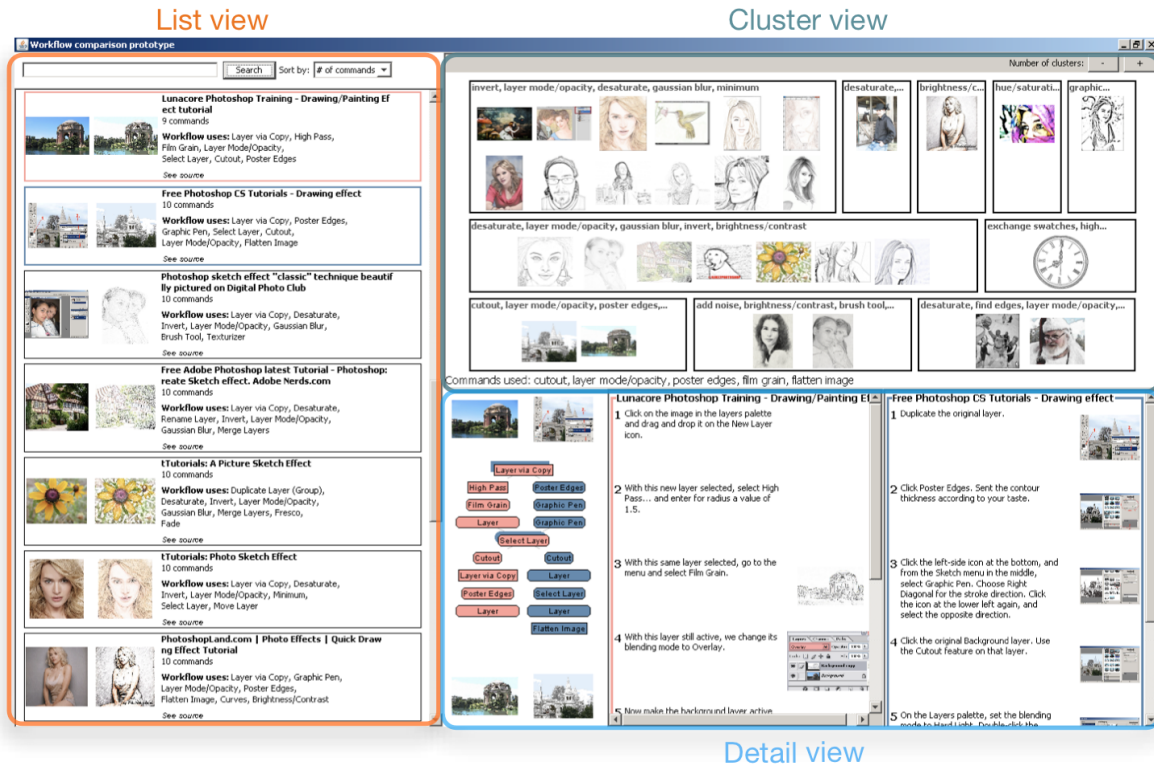
List view   Cluster view

Detail view

Figure 3. Delta's workflow explorer. The list view (left) displays workflow summaries. The cluster view (top right) groups workflows by similarity. The detail view (bottom right) shows individual steps and *union graphs* of shared commands.

Delta labels each cluster with the five most common tools in the cluster's workflows. Labels may be abbreviated because of cluster sizes; the full label is displayed at the bottom of the view when users mouse over a cluster. Users can click on an image in the cluster view to cause the list view to scroll to the corresponding workflow. Finally, users can adjust the number of clusters using *plus* and *minus* icons to explore coarser or finer corpus differences.
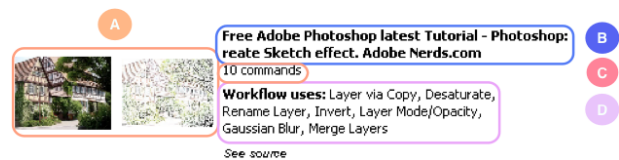
List View

The list view (Figure 3 left) uses **workflow summaries** (Figure 4) to surface the most salient attributes we found in Study 1. Reading an abstract summary can enable quicker comparison than consulting the source documents.

Summaries show *result quality* by displaying the before and after thumbnail images and a title. Users can click on thumbnails to view a full-resolution version. Summaries enable users to judge *command familiarity* through a list of unique commands in the workflow. The summary also displays the total number of commands, which is a coarse indicator of *efficiency*. Each workflow view also contains a link ("*See source*") to the source webpage. Users can sort the list view by either the number of commands or by cluster (which orders workflows so siblings in a cluster are adjacent).

Detail View

The detail view shows individual steps in a workflow and also enables pairwise comparison. When a user selects a

Figure 4. A workflow summary. (a) Before and after images. (b) Workflow title. (c) Number of commands. (d) List of unique commands in the workflow.
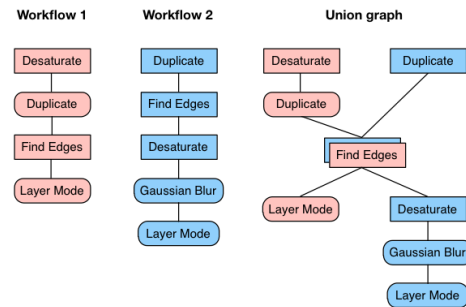

Figure 5. Creation of the union graph. Only commands that share the same name and parameters overlap. Here, "Find Edges" is shared between the two workflows.

single workflow in the list view, the detail view displays all commands and any accompanying screenshots of the original tutorial; the layout was modeled after the *text+image* condition in Study 1. When the user has selected two workflows, the system displays the workflows

side-by-side and creates a *union graph* [1] of the two command sequences. Figure 5 illustrates the creation of the union graph, which is produced using the edit distance computed for the cluster view. Commands in both workflows that have the same *name* and *parameters* are considered to be shared between workflows and are visually centered in the graph; in Figure 5, "Find edges" is the only shared command. Commands without parameters are displayed as rectangles and commands with parameters are displayed as rounded rectangles. Finally, node tooltips show the full name of the command and any parameters for that command. The graph allows the user to judge similarity by identifying shared commands.

STUDY 2: WORKING WITH DELTA

We designed a study to evaluate how effectively the different views of our system were able to aid users in exploring a small corpus of workflows. We also compared Delta to web search in order to contrast usage behaviors. However, the primary purpose of the study was to observe user behavior with Delta and gain insights into potential design enhancements.

Methodology

We asked users to complete a set of workflow comparison and search tasks on a corpus of 30 similar workflows while thinking-aloud. Specifically, we asked users to tell us which features of the workflows or interface they were attending to while completing the tasks.

To compare user behavior in Delta with a baseline, we tested two conditions: *Delta* – where users worked with our prototype workflow explorer tool, and *WebSearch* – where participants used a web search interface over the study corpus[1]. In each interface condition, participants saw a different set of workflows: one for *image sharpening*, and one for *applying a sketch effect*. Each corpus contained 30 workflows, which we manually transcribed from tutorials taken from the top Google search results. The source workflows were of similar length and scope as the workflows we presented to users in Study 1. We found that 30 workflows gave us a good variety of approaches, although the small corpus may have limited the utility of our system.

The ordered tasks we gave participants were as follows (phrased using the sketch effect task):

1. Find a workflow you would use to apply a sketch effect. (*Find a workflow*)

2. You have an image that you would like to apply a sketch effect to, but you are not sure which method would produce the best result. Find three workflows that each use a different approach to apply a sketch effect. (*Three approaches*)

3. You were following this sketch effect workflow and got stuck midway because one step was not clearly explained. Find the workflow most similar to this workflow. (*Most similar*)

4. You are nearing a deadline and don't have much time. Find the sketch effect workflow that would be quickest to complete. (*Quickest*)

5. You have some time and would like to learn more about Photoshop. Find the sketch effect workflow that would expose you to the most new or unfamiliar tools. (*Most new commands*)

We call tasks 2 and 3 **comparison tasks**, and the other tasks **search tasks**.

Finally, we asked participants to complete a questionnaire asking (a) what parts of Delta they liked or disliked, and (b) which interface they preferred in each task.

Participants

We used an online classified posting to recruit participants, asking for people who had at least beginner-level Photoshop experience. Twelve participants (six male and six female) completed the study. They ranged in age from 26-67 years old and reported between two months and 20 years of Photoshop experience.

Results

We first discuss our observations of user behavior when using Delta, then briefly report user ratings of their preferred interface for each task. We do not report time-to-completion for either interface as the corpus was small enough that users tended to learn the workflows quickly, which meant that they were sometimes able to complete the later tasks very quickly by referencing a workflow they had already seen.

Participants Used the List View for All Tasks

When using Delta, we noted that users depended heavily on the list view for all of the tasks. All users quickly completed the search tasks by taking advantage of the list view's sort by number of commands, from fewest to most. For the *Quickest* task, users scrolled to the top of the list and chose one of the top workflows based on the result quality and their own familiarity with the commands. In the *Most new commands* task, users scrolled to the bottom of the list to find the workflows with the most commands, then read the list of commands in the workflow summaries and chose the one that contained the most unfamiliar commands.

In the comparison tasks, all users made use of the command list in the workflow summaries as their primary way of finding candidate workflows. In fact, P9 exclusively used the list view for both tasks. He rapidly judged similarity from the command list; for example, in the *Most similar* task, he said, *"both [workflows] have similar tools"* and did not look at the details of the workflows.

---

[1] We implemented this condition using the Google Custom Search API (http://code.google.com/apis/customsearch/).

Participants Largely Ignored the Cluster View
On the other hand, only three users interacted with clusters, either in the cluster view or by reordering the list view by cluster. P3 clicked once in the cluster view during the *Three approaches* task and sorted by cluster in the *Most new commands* task. P7 reordered the list view by cluster during the *Three approaches* task, but relied on the content of the source images to choose three different workflows. P4 used the cluster view in the *Most similar* task to choose a workflow in the same cluster as the supplied workflow. P4 then selected the workflow in the same cluster that had the most command overlap.

Most Participants Used the Detail View and Union Graph
Four of the twelve participants used the detail view in the *Find a workflow* task, and ten participants used the detail view to read individual steps in the workflow when completing the comparison tasks. They would first use the list view to find interesting workflows, then select the workflow and read the details.

Seven of the twelve participants used the union graph in at least one of the comparison tasks. P11 commented that, *"A lot of the commands are joined"* while completing the *Most similar* task. Others, such as P2, noted shared commands, but also commands that are shared but out of order:

"The only [command] they share is layer via copy … and some of their directions are pretty much reversed. One has a layer [blending] after a layer via copy, the other has a layer [blending] after another command."

This observation indicates that users understood the concept of the union graph and found it useful for completing the comparison tasks. It also suggests that explicitly indicating shared, but out of order, commands is potentially useful.

Although users understood the main concept of the union graph, there were some features that caused confusion. Some users did not recall that hovering on a command in the graph activated a tooltip with the full name of the command and its parameters; when referencing a command, participants sometimes used the shortened display name of the command rather than the actual command. A number of users were also confused that commands with the same name were not joined in the union graph unless they also shared the same parameters. None of the participants used the parameters of commands when choosing workflows; this implies this information does not need to be present in the views, as users can retrieve the information from the source document.

Participants Used Images Extensively
All of the participants viewed the full resolution versions of the before or after images to compare the results, and many users viewed the full resolution images that illustrated each step. However, although our design allowed users to pop open multiple full-resolution images to aid comparison, only P11 moved image windows around to compare a full-resolution image to another image. These observations
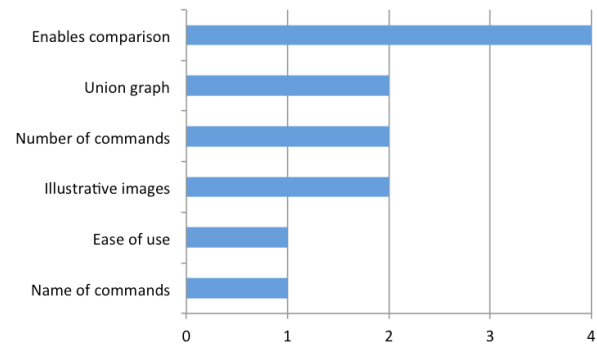

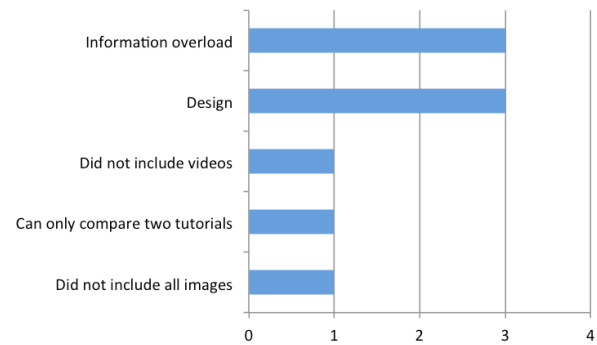**Figure 6. Elements of the workflow explorer that users reported they liked.**

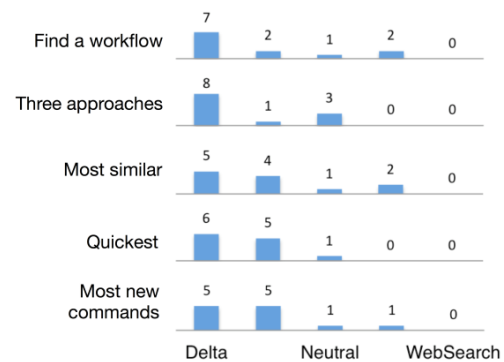**Figure 7. Elements of the workflow explorer that users reported they disliked.**

Figure 8. User interface preferences for each task. Numbers indicate the number of ratings.

suggest that while full-resolution images were important for users to see the details of the results, the thumbnail images were good enough for users to compare workflows.

Subjective Feedback
We asked users to tell us what they liked and disliked about Delta. We categorized their open-ended responses: Figure 6 shows what users liked about the application and Figure 7 shows what users disliked about the application. Overall, users liked how the application enabled them to compare workflows, and two users mentioned the union graph in particular, both referring to the graph as a "*flow chart*." The

main drawback, noted by some of the participants, was that the interface was filled with too much information. Finally, participants consistently preferred Delta over the search engine for each task (Figure 8).

Follow-Up Study: Clustering Methods

We also wanted to learn about how users would cluster workflows, to better understand what areas of the similarity metric space would be useful. We gave the same twelve participants from our user study 10 printed workflows for red-eye correction and asked them to create groups of workflows that use *similar approaches* to complete the task. The printed workflows were in the *text+image* format from Study 1. We intentionally left the prompt vague to allow participants to use their own notion of similarity. We asked participants to think-aloud while grouping workflows, and specifically asked them to tell us why they placed workflows in the same group. The clustering study took approximately 10 minutes to complete.

Results

We noticed two primary methods of clustering: using commands, or command groups, and using the complexity of the workflows (based on familiarity, length, or ease of the workflow). Some users combined these two methods when creating their clusters.

Seven users used command groups to group workflows. For example, P1 separated the workflows into one group which involved "*filling in the eye with different colors*", one that "*changed colors using channels*", and one that involved "*painting on new layers.*" A few made a distinction between workflows that used layers versus those that used painting tools, such as P9.

Five users clustered workflows based on complexity. A few clustered based on how easy the workflow was to follow; for example, P5 created one cluster with workflows that had "*too much information,*" and one cluster with workflows that were "*nice and simple.*" P12 split workflows into familiar and unfamiliar groups, and further clustered the familiar group using commands.

These results indicate that individual users may group commands differently, and use those groups to judge similarity. In addition, users thought it was important to differentiate between easy and difficult workflows. The implication for design is that an effective clustering metric would combine both command similarity and a measure of workflow complexity.

DISCUSSION

Implications of Observations: Design Guidelines

The observations from our user study imply a number of design guidelines for a future workflow exploration system.

Provide List Views to Support Search Tasks

Participants universally used the list view for all tasks, and largely ignored the cluster view. Based on these results, it is likely that the list view contained the required information to complete the tasks, given the small size of the workflow

corpus. The cluster view may have been unnecessarily high level. We hypothesize that as the number of workflows increases, the cluster view may provide more benefit. The larger the corpus, the more beneficial a high-level overview of the collection is likely to be, as prior clustering systems have shown [7, 23].

Provide Workflow Details to Support Comparison Tasks

In comparison tasks, users used the list view to identify interesting workflows, then read the details of the workflow to make a final judgment. Although one user only used the workflow summaries to complete the comparison tasks, the others read the details of the workflow. This indicates that providing access to the source material is imperative.

Show One View at a Time to Reduce Information Overload

One of the biggest problems users encountered was information overload; in fact, one user did not complete the *Three approaches* task, saying he was "overwhelmed." A possible solution lies in our observation that users did not rapidly switch between views. In addition, although the list view and cluster view were linked (clicking on a workflow in the cluster view scrolled to the corresponding workflow in the list view), users did not make use of this link. This means that the system could display panels separately instead of all at once on the screen.

Study Limitations

Our study has several limitations worth noting. Participants were introduced to Delta through a five minute walkthrough of the interface by the experimenter. Given this short amount of time, users may have preferred familiar interface components, such as the list view, at the expense of unfamiliar components, such as the cluster view. Extended use with the system may result in different behavior.

In Study 1, users reported that they used tutorials serially: they would find a workflow, try it out in Photoshop, and only look for another if the first wasn't satisfactory. Users might exhibit different behavior if our system were integrated into Photoshop; we leave this to future work.

Finally, we did not test how accurate users' choices were. For example, we do not know if users chose the workflow that would be quickest for them to complete, as we did not have them complete the workflow after they had chosen it. However, our tool was designed to elicit user behavior and was not designed as a final solution to the problem of workflow exploration.

FUTURE WORK AND CONCLUSION

We have studied users' criteria when choosing workflows and designed a new interactive application to support that process based on the results. Our study indicates that the application provides subjective benefits in comparison to the traditional means of using a web browser to locate and compare workflows.

The results of our studies provide many directions for future work. In particular, we found that users consider command familiarity when comparing workflows.

Integrating this tool into a system would allow us to use an individual's command use to provide better ordering in the list view, or a better similarity metric for the cluster view.

In addition, we did not use any "broken" workflows in our study, or workflows that do not result in their claimed effect. Testing whether our system allows users to avoid such workflows is another direction for future work.

We would also like to explore how well our system scales and what designs will work best as the number of workflows increases, or when the workflows become longer, such as those in Chronicle [11] or MeshFlow [8].

Finally, we are actively working on automatically extracting workflows from websites that contain tutorials. Our results indicate that just the accurate extraction of commands and illustrative images from a tutorial would greatly enhance search efficacy. In addition, automatic workflow extraction will help us explore how our tool scales by opening our tool to a large number of workflows. Although using workflow recording systems would be an easy method to provide data to our system, such systems are currently not widely used.

ACKNOWLEDGMENTS

REFERENCES

1. Andrews, K., Wohlfahrt, M., and Wurzinger, G. Visual Graph Comparison. *Info. Vis.,* pp. 62-67, 2009.

2. Bederson, B., Shneiderman, B., and Wattenberg, M. Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies. *ACM Trans. Graph*, **21**(4):833-854, 2002.

3. Chen, H.-T., Wei, L.-Y. and Chang, C.-F. Nonlinear Revision Control for Images. *ACM Trans. Graph*, **30**(4), Article 105, 2011.

4. Chi, E. H., Pirolli, P., Chen, K., and Pitkow, J. Using information scent to model user information needs and actions and the Web. *ACM CHI,* pp. 490-497, 2001.

5. Chi, E., Pirolli, P., and Pitkow, J. The scent of a site: a system for analyzing and predicting information scent, usage, and usability of a Web site. *ACM CHI*, pp. 161-168, 2000.

6. Cutrell, E. and Guan Z. What are you looking for?: an eye-tracking study of information usage in web search. *ACM CHI*. pp. 407-416. 2007.

7. Cutting, D., Karger, D., Pedersen, J., and Tukey, J. Scatter/Gather: a cluster-based approach to browsing large document collections. *SIGIR*. pp. 318-329. 1992.

8. Denning, J. D., Kerr, W. B., and Pellacini, F. MeshFlow: Interactive Visualization of Mesh Construction Sequences. *ACM Trans. Graph*, **30**(4), Article 66, 2011.

9. Dorn, B. and Guzdial, M. Learning on the job: characterizing the programming knowledge and learning strategies of web designers. *ACM CHI,* pp. 703-712. 2010.

10. Grabler, F., Agrawala, M., Li, W., Dontcheva, M. and Igarashi, T. Generating Photo Manipulation Tutorials by Demonstration. *SIGGRAPH*. pp. 66:1-66:9, 2009.

11. Grossman, T., Matejka, J. and Fitzmaurice, G. Chronicle: Capture, Exploration, and Playback of Document Workflow Histories. *UIST,* pp. 143-152, 2010.

12. Grossman, T., Fitzmaurice, G. and Attar, R. A Survey of Software Learnability: Metrics, Methodologies and Guidelines. *ACM CHI*, pp. 649-658, 2009.

13. Guan. Z., and Cutrell, E. An eye tracking study of the effect of target rank on web search. *ACM CHI*, pp. 417-420, 2007.

14. Hearst, M. *Search User Interfaces* (1st ed.). Cambridge University Press, New York, NY, USA. 2009.

15. Heer, J., Mackinlay, J., Stolte, C. and Agrawala, M. Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation. *IEEE TVCG,* pp. 1189-1196, 2008.

16. Jing, F. Wang, W., Yao, Y., Deng, K., Zhang, L., and Ma, W.-Y. IGroup: web image search results clustering. *ACM Multimedia*, pp. 377-384, 2006.

17. Koop, D., Scheidegger, C., Callahan, S., Freire, J., and Silva, C. VisComplete: Automating Suggestions for Visualization Pipelines. *IEEE TVCG*, pp. 1691-1698, 2008.

18. Kurlander, D.,and Feiner, S. Editable graphical histories. *IEEE Visual Languages, pp.* 127-134, 1988.

19. LaFreniere, B., Bunt, A., Lount, M., Krynicki, F., and Terry, M. AdaptableGIMP: designing a socially-adaptable interface. *UIST Adjunct*, pp. 89-90. 2011.

20. Lee, B., Srivastava, S., Kumar, R., Brafman, R., and Klemmer, S. Designing with Interactive Example Galleries. *ACM CHI*, pp. 2257-2266, 2010.

21. Nakamura, T. and Igarashi, I. An application-independent system for visualizing user operation history. *UIST*, pp. 23-32, 2008.

22. Pirolli, P. and Card, S. Information foraging. *Psychological Review*, **106**(4):643-675, 1999.

23. Santos, E., Lins, L., Ahrens, J., Freire, J., and Silva, C. A first study on clustering collections of workflow graphs. *IPAW*, pp. 160-173, 2008.

24. Sorokin, A. and Forsyth, D. Utility data annotation with Amazon Mechanical Turk. *CVPRW*, pp.1-8, 2008.

25. Wagner, R. and Fischer, M. The String-to-String Correction Problem. *J. ACM,* **21**(1):168-173, 1974.