

# How Do People Organize Their Photos in Each Event and How Does It Affect Storytelling, Searching and Interpretation Tasks?

Jesse Prabawa Gozali<sup>1</sup>   Min-Yen Kan<sup>1</sup>   Hari Sundaram<sup>2</sup>  
<sup>1</sup>Department of Computer Science, National University of Singapore, Singapore  
<sup>2</sup>Arts Media & Engineering, Arizona State University, USA  
{jprabawa, kanmy}@comp.nus.edu.sg   hari.sundaram@asu.edu

## ABSTRACT

This paper explores photo organization within an *event photo stream*, *i.e.* the chronological sequence of photos from a single event. The problem is important: with the advent of inexpensive, easy-to-use photo capture devices, people can take a large number of photos per event. A family trip, for example, may include hundreds of photos. In this work, we have developed a photo browser that uses automatically segmented groups of photos—referred to as chapters—to organize such photos. The photo browser also affords users with a drag-and-drop interface to refine the chapter groupings.

We conducted an exploratory study of 23 college students with their 8096 personal photos from 92 events, to understand the role of different spatial organization strategies in our chapter-based photo browser, in performing storytelling, photo search and photo set interpretation tasks. We also report novel insights on how the subjects organized their photos into chapters. We tested three layout strategies: bi-level, grid-stacking and space-filling, against a baseline plain grid layout. We found that subjects value the chronological order of the chapters more than maximizing screen space usage and that they value chapter consistency more than the chronological order of the photos. For automatic chapter groupings, having low chapter boundary misses is more important than having low chapter boundary false alarms; the choice of chapter criteria and granularity for chapter groupings are very subjective; and subjects found that chapter-based photo organization helps in all three tasks of the user study. Users preferred the chapter-based layout strategies to the baseline at a statistically significant level, with the grid-stacking strategy preferred the most.

## Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL'12, June 10–14, 2012, Washington, DC, USA.

Copyright 2012 ACM 978-1-4503-1154-0/12/06 ...\$10.00.

## General Terms

Design, Human Factors.

## Keywords

Photo browser, photo digital library, photo layouts, event photo stream segmentation

## 1. INTRODUCTION

Today, people take more photos—with the help of inexpensive, easy-to-use and portable photo capture devices—in contrast to an earlier era of film rolls and analog cameras [17]. Not surprisingly, large photo collections are becoming more common. People with large collections are everyday photographers, interested in simply capturing the moment, in contrast to the serious amateur or professional photographer of the analog era. Most personal photos are commonly associated with an event: a holiday trip, picnic, dinner or walk in the park. Many academic and commercial photo browsers, like iPhoto<sup>1</sup> and Picasa<sup>2</sup>, advocate event-based photo organization. With the ease of photo-capture however, a family trip can contain hundreds of photos—sifting through these event photos is still cumbersome.

To complement event-based photo organization and help make photos of each event more manageable, we propose to organize photos in each event into smaller groups of photos, *i.e.* all of the groups belong to *the same event*. Figure 1 shows an example *event photo stream*, where the chronological sequence of photos from a single event is segmented to produce groups of photos, each corresponding to a photo-worthy moment in the event. This paper explores photo organization within such an event photo stream. We want to answer the following question: *How do people organize their photos in each event and how does it affect storytelling, searching and interpretation tasks?*

We have developed a photo browser called *Chaptrs* that helps users organize their event photos by automatically grouping photos in each event into smaller groups of photos we call chapters. *Chaptrs* builds upon our prior work to segment an event photo stream [?]. The *Chaptrs* photo browser also affords users with a drag-and-drop interface to refine the chapter groupings.

With *Chaptrs*, we conducted an exploratory study involving 23 college students with a total of 8096 personal photos

<sup>1</sup><http://www.apple.com/ilife/iphoto>

<sup>2</sup><http://picasa.google.com>



**Figure 1: Part of an event photo stream is segmented into smaller groups of photos. We refer to each group as a chapter of the event. The chapters are labelled in this example.**

from 92 events. To facilitate the study, we implemented four photo layouts in *Chaptrs* (see Figures 2, 3, 4, and 5). The first is our baseline, a plain grid layout commonly used by commercial photo browsers and offers no chapter-based photo organization. The other three layouts present chapter-based photo organizations but each emphasizes on a different key aspect. The **bi-level layout** emphasizes an overview of the event photos afforded by presenting chapter thumbnails. The **grid-stacking layout** emphasizes the chronological order of the chapters. Lastly, the **space-filling layout** maximizes screen space usage.

The three chapter-based layouts were chosen because they emphasize and represent distinct key layout aspects. As such, they facilitated our study to explore which key aspects are important for chapter-based photo organization. To our knowledge, our study is the first to explore chapter-based photo organization and its photo layouts.

The key research contributions in this paper are as follows:

- *Novel insights on how users group their event photos into chapters.* The main insights are as follows: users value chapter consistency more than the chronological order of the photos; criteria for chapters include moment, object, location, photography type, and intention; and that the choice of criteria and granularity for segmentation are very subjective.
- *Explore how chapter-based photo organization affects three key photo-related tasks.* We found several statistically significant results, among them: subjects found grouping event photos by chapter helps present the event’s story, helps to find a photo, and helps to interpret unfamiliar photos. In contrast, the plain grid layout without chapter-based photo organization was preferred the least for all three tasks.
- *Explore key photo layout aspects for chapter-based photo organization.* Most subjects value the chronological order of the chapters more than maximizing screen space usage. Subjects also appreciate having an overview of the event photos afforded by chapter thumbnails.

In the next section, we review related works on photo organization and photo layouts in personal digital photo libraries. Section 3 describes the photo layouts. We present the user study and its results in Sections 4 and 5. Finally, we conclude in Section 6.

## 2. RELATED WORK

### 2.1 Photo Organization

Over the past decade, there have been a number of studies on how people manage—including organization and sharing—their personal photo collections. Rodden [24, 25] has studied how people manage their photo collections, printed or otherwise. Some findings from his study include: printed photo

albums are mostly classified by event, with one album for each event. Searching a printed photo collection is typically done for a photo album of a specific event. Even if the search was for a specific photo, people will try to locate the album containing the photo first before starting the search. For personal digital photo libraries, people regard the ability to organize photos into folders as very useful and would arrange them according to events in a chronological order. People prefer to browse their photos by event rather than querying. Similar findings were also found by Cunningham and Masoodian [7]. They conclude that browsing, rather than searching, is a more practical tool for locating photos.

Other studies go beyond how the photos are organized. Kirk *et al.* [17] coined the term “photowork”, *i.e.* activities done after photo capture but before sharing. These include reviewing, downloading, organizing, editing, sorting, as well as filing of photos. Frohlich *et al.* [10] conducted a study to establish requirements for photo sharing technologies. A recent article by Sandhaus and Boll [27] presents a good overview of research in this field of personal photo collections. To our knowledge, however, our study is the first to explore chapter-based photo organization.

### 2.2 Photo Layouts in Personal Digital Photo Libraries

An effective photo layout is one that presents photos in a way that supports users in one or more photo-related tasks. Here, we review existing works on photo layouts for personal digital photo libraries to gather the key aspects they emphasize and the tasks they support effectively.

While there has been prior work to study event-based photo organization, the absence of work on photo layouts for chapter-based photo organization, *i.e.* layouts to present groups of photos with all groups belonging to *the same event* is notable. In event-based photo organization, the groups of photos belong to *different events*. The closest work we found was by Graham *et al.* [15]. They proposed a hierarchical photo browser to better support search tasks by presenting a 25 photo summary at various levels of hierarchy of the user’s photo collection: year, month, event, and also for groups of photos within an event. The user navigates through the view hierarchy using a tree view in the sidebar.

For event-based photo organization, the most common photo layout is a 2D grid: photos are ordered chronologically row by row on a grid. Many photo browsers [18, 20, 8, 19] including commercial ones like Picasa and iPhoto adopt this layout to display photos of an event. A plain grid layout is a simple layout that maximizes use of the available screen space. Having many photos visible at once allows users familiar with the photos to scan them very quickly [25].

Photo browsers typically display one event (one grid) at a time, but some photo browsers relieve users from having to select individual events from the view hierarchy by displaying all the events at once: the grids are stacked on top

of each other in chronological order, *e.g.* Picasa. The layout remains uniform as the grids have the same number of columns. With this layout, users can browse their events by simply scrolling. To demarcate the events, each grid has a title bar on top with the event information. Alternatively, in the timeline view of [20], each grid is labeled hierarchically on its left margin by month and year. In [4], all the photos in the collection are displayed as one massive grid and event titles are displayed as grid elements to demarcate the events.

Time Quilt[16], a zoomable photo browser designed to enhance search tasks, also displays photos from all events at once. Its layout trades-off screen space usage for better presentation of the chronological order of the photos. Photos from each event are displayed in their own grid. The grids are then displayed chronologically column by column. The number of rows and columns of each grid follows the aspect ratio of the corresponding thumbnail of the event. Each grid is replaced with the event thumbnail of the same size and the grid only becomes visible when the user zooms in.

Some photo browsers do not use a grid layout. Tree-Browser[5] is a photo browser for multiple photo collections. The collections are displayed chronologically at the top of the photo browser as a single scrollable row of thumbnails. The main part of the photo browser displays events from the selected collection as a tree of depth one. The tree root is the collection thumbnail. Each leaf corresponds to an event in the collection and is displayed as a single row of photos.

The photo browsers we have reviewed so far have layouts that emphasize one or more of the following key aspects: use of view hierarchy, chronological order of event photos, and maximization of screen space usage. We emphasize on similar key aspects in the three layouts used for the user study: the bi-level, grid-stacking and space-filling layouts.

The works we have reviewed have also weaved the chronological order of the photos into two dimensions (*e.g.* row-by-row) to make better use of screen space. However, in interfaces where visualizing the timeline is more important, chronological order is commonly conveyed as a single dimension in the layout [23, 9, 1]. Photo storytelling interfaces exhibit similar linear structures in their layouts. Here, we highlight three notable interfaces: the first two are well-cited and the third is a recent contribution to the field. First is the story-editing environment in FotoFile[18]. Here, users can select photos from an *Image Tape* at the top of the photo browser and place them into one of the row of *Scraplets* in the main part of the photo browser. Each scraplet displays its photos as a single column. Balabanović *et al.* [2] developed a portable device for sharing and authoring stories. In its interface, the navigation area consists of rows of photo thumbnails. Photos in the rows are shown in groups of alternating backgrounds to distinguish separate photo rolls. Recently, Raconteur[6] is a story editing system that helps users assemble stories from annotated media files. The media files are arranged in chronological order in a single row.

Some photo browsers were designed with the key aspect of effectively conveying inter-photo similarity, *e.g.* in terms of visual appearance, location, or tag. These photo browsers generally present more visually interesting and novel layouts. However, the chronological order of the photos often suffers as a result. For example, PhotoMesa[3] employs quantum treemaps and bubblemaps to display labelled photo clusters in a grid layout to maximize screen space usage. More recently, MediaGlow[12] uses a spring layout algorithm to help



Figure 2: Plain grid layout

users stack and retrieve similar photos. PHOTOLAND[26] presents a layout that places photos on a 2D grid based on an inter-photo similarity measure computed from temporal and spatial information. The result is a layout that presents photos from an event as an island of thumbnails.

Following the motivation behind event-based photo organization, we suspect that similarly, users will create chapters corresponding to moments in the event. As such, the chapter-based layouts in *Chaptrs* present the chronological order of the chapters in different ways. Following the works we have reviewed here, another aspect to explore is the issue of screen space usage maximization. As such, the chapter-based layouts in *Chaptrs* have varying degrees of screen space utilization. Lastly, the bi-level layout presents a two-level view hierarchy to provide an overview of all the event photos using a horizontal film strip of chapter thumbnails.

### 3. PHOTO LAYOUTS USED FOR STUDY

In *Chaptrs*, we implemented four layouts for displaying photos from a single event (see Figures 2, 3, 4, and 5):

1. **Plain grid layout** is our baseline layout and it consists of a single grid of row-by-row chronologically-ordered photos. No chapter information is presented in this layout.
2. **Bi-level layout** consists of a split view where the bottom view displays a film strip of chronologically-ordered chapter thumbnails for selection and the top view displays photos of the selected chapter in a grid layout, in chronological order row-by-row.
3. **Grid-stacking layout** consists of chronologically-ordered vertically-stacked grids, each corresponding to a chapter. Photos in each grid are ordered chronologically row-by-row.
4. **Space-filling layout** consists of a single grid of row-by-row chronologically-ordered event photos with an outline surrounding each span of photos that are part of the same chapter.

*Chaptrs* also affords users with a drag-and-drop interface to edit the chapter groupings in the bi-level layout. By default, our event photo stream segmentation algorithm automatically groups event photos into chapters so users only



Figure 3: Bi-level layout

need to adjust the chapter groupings instead of starting from scratch. To combine adjacent chapters, users simply drag one chapter thumbnail onto another from the film strip. When users have a chapter selected in the film strip, its photos are shown in the top view. To move photos into a new chapter, users can select a span of photos at the beginning or end of the chapter and then drag the photos onto the film strip. Other kinds of selections are not valid to ensure that the chronological order of the photos in the stream is not violated. The four layouts take inspiration from our review of existing photo layouts for personal digital photo libraries. We adapt them to organize chapters, instead of other group types (*e.g.* events, similar photos). The bi-level layout takes inspiration from photo storytelling interfaces which present the chronological order unweaved in a single horizontal dimension. The space-filling layout takes inspiration from the bubblemap layout in PhotoMesa and maximizes screen space usage. The grid-stacking layout is similar to how Picasa displays photos from all events at once with a separate grid for each event. Screen space is still wasted but not as much as in the bi-level layout. We now discuss each of the chapter-based layouts in more detail.

### 3.1 Bi-Level Layout

The bi-level layout consists of a split view where the bottom view provides an overview of all the photos by displaying a scrollable film strip of chapter thumbnails. The top view displays photos from the selected chapter in a grid layout.

Chapter thumbnails are displayed in chronological order. Each thumbnail is labelled with the timestamp of the first photo in the corresponding chapter and optionally, labelled with a user-defined title. The film strip provides users with an overview of all the photos. It acts as an index into the event photos, allowing users to glean over moments in the event through the chapter thumbnails without having to sift through individual photos. The chapter groupings allow users to collapse the timeline in a meaningful way and present chapter thumbnails in a linear structure that effectively conveys their chronological order.

### 3.2 Grid-Stacking Layout

The grid-stacking layout displays all photos from the event with photos of each chapter in its own grid. Photos in each

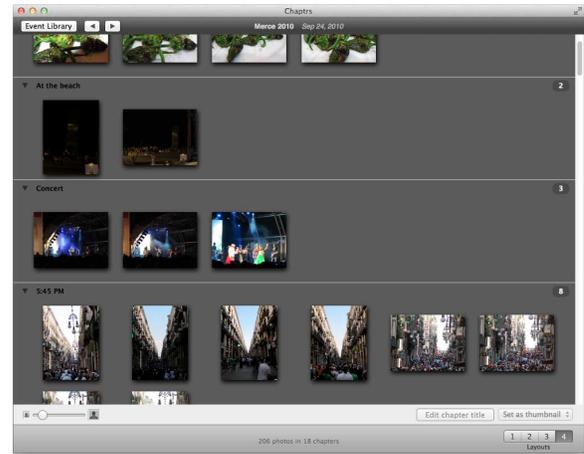


Figure 4: Grid-stacking layout

grid are ordered chronologically row-by-row. All grids have the same number of columns and are displayed in chronological order separated by a horizontal line and chapter title.

Compared to the bi-level layout, the grid-stacking layout makes better use of screen space. While the grids may not be fully occupied with photos, the grids are stacked one after another, leaving no room between adjacent grids. The chronological order of the chapters are also presented in a linear structure by stacking the grids in one dimension.

### 3.3 Space-Filling Layout

The space-filling layout displays all photos from the event in a single grid. Photos are ordered chronologically row-by-row. In addition, an outline is drawn around photos of the same chapter. To keep photos contiguous within each chapter outline, some grid elements may be left empty (see Figure 5). This layout is similar to the bubblemap layout in PhotoMesa but maintains a row-by-row chronological order. As such, the space-filling layout is not as densely packed and may still waste some screen space.

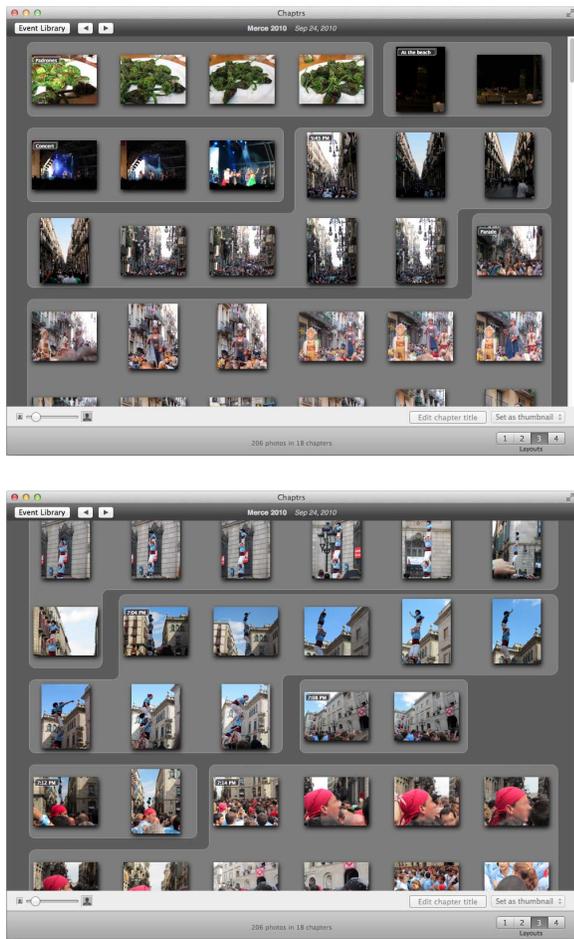
Of the three chapter-based layouts, the space-filling layout is the one that wastes the least amount of screen space and displays the most number of thumbnails at once while still presenting the chapter groupings. These space savings are however, at the expense of the chronological order of the chapters. Unlike the grid-stacking layout, the chronological order of the chapters is weaved into two dimensions row-by-row, instead of linearly top-down.

## 4. USER STUDY

The goal of our user study is to explore three questions:

1. How do people organize their photos in each event?
2. How does chapter-based photo organization affect photo-related tasks such as storytelling, searching, and interpretation tasks?
3. What photo layout aspects are important for chapter-based photo organization?

We recruited 23 college students. One subject is a professional photographer who often participates in photography trips at public events or at leisure. Another subject maintains an active food blog and always has a digital camera



**Figure 5: Space-filling layout: Event photos are displayed in a grid layout, in chronological order row-by-row, with an outline surrounding photos of the same chapter. The bottom figure illustrates how some grid elements may be left empty in order to keep photos contiguous within each chapter outline.**

at hand. Some are enthusiastic amateur photographers who carry their digital cameras for social events. Others only carry their digital cameras during holiday trips. Most subjects use Windows Explorer or Windows Live Photo Gallery as their primary photo browser. Some use Picasa, two subjects use iPhoto, and one subject uses Aperture.

Following Institutional Review Board exemption guidelines, photos were immediately discarded at the end of each study session and all collected data was anonymized.

## 4.1 Photo Sets

Subjects were asked to bring four sets of personal photos, each from a different event. While most events are associated with holiday trips, others span a variety of event types: a public cosplay event, a college orientation camp, talks at a conference, a stage performance, visit to the museum, etc. The total number of photos in the study is 8096 photos from 92 photo sets. We asked the subjects to bring at least one set with more than 100 photos and at least one with 40-60 photos. This allowed us to ask the subjects to reflect on sets with many photos or few photos. Before we imported

the subject’s photo sets into *Chapters*, we asked the subject to choose four different favorite photos from the set with the most photos, using the default file explorer application. These photos were later used in the searching task.

After the photo sets have been imported, we asked the subject to “group the photos into chapters according to their preference and liking”. Additionally, we randomly selected two photo sets from the subject for s/he to group into chapters *without help* from our event photo stream segmentation algorithm, *i.e.* the subject started with *no initial chapter groupings*. For his/her photo sets, we asked the subject to group the photos to his/her satisfaction; the subject’s final organization for the photo sets is used for the study tasks. This protocol allowed us to analyze the effects of initializing the chapter groupings on how the subjects group their photos into chapters.

## 4.2 Study Tasks

Subjects were asked to complete three tasks. Subjects were also asked to fill a questionnaire after each task, and another overall questionnaire after all three tasks. All questionnaires use a standard 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Finally, each study session ended with a semi-structured interview<sup>3</sup>. The audio from the interview session was recorded for note-taking purposes.

In our study, we focused on common photo-related tasks for users – tasks that fit the STU (Situations, Tasks, and Users) context [22]. In particular, the first two tasks have been used in the related works we reviewed in Section 2. We describe each task in more detail next, followed by more details on how we eliminated confounding variables.

### 4.2.1 Task 1: Storytelling from familiar event photos

In this task, subjects were asked to tell the story of each event from their personal photo sets. We asked subjects to imagine sharing about the event and its photos, as they normally would, to their friends. We used a within-subject design where each subject carries out the task four times, each with a different layout. To avoid learning effect on the story told, each layout was used with a different photo set.

### 4.2.2 Task 2: Finding a given photo from familiar event photos

In this task, subjects were asked to find the favorite photos they chose at the beginning of the study. We used a within-subject design where each subject carries out the task four times, each for a different favorite photo and with a different layout. At each iteration, the target favorite photo was clearly displayed on an adjacent external monitor. The four favorite photos were chosen from the same photo set to make the iterations comparable. There is no learning effect between iterations on the photo set because the subject – who also owns the photo set – has been through the photos at least twice from the storytelling task and from grouping the photos into chapters at the beginning of the study.

### 4.2.3 Task 3: Interpreting unfamiliar event photos

In this task, subjects were shown and asked to interpret unfamiliar event photos. We asked the subjects: “Tell me about the event. What do you think was happening?”. For this task, we prepared four sets of event photos that were not used in any other part of the study. The photo sets

<sup>3</sup>Questionnaires and interview questions available in [14]

were titled, grouped into chapters, but chapters were left untitled. We used a within-subject design where each subject carries out the task four times, each with a different layout. To avoid any learning effects, each layout was used with a different photo set. This task is the most synthetic of the three tasks in our user study. While subjects are unlikely to find themselves having to interpret event photos without any context other than the photos themselves and the event title, our goal was simply to create a scenario where the subjects have very little knowledge of the event, similar to how they would find themselves when faced with an old set of event photos but not remembering any details of the event [10].

### 4.3 Internal Validity

We chose a within-subject design, *i.e.* repeated measurements per subject, to have better internal validity, as is common for user studies with few subjects. The personal nature of the photos and the length of the study per subject made recruiting hundreds of subjects impractical.

As mentioned in Section 4.2, we have tried to eliminate any learning effects. In addition, we eliminated learning effects on the four layouts by demonstrating *Chaptrs*, its four layouts, and all their features at the beginning of the study, prior to any of the tasks. We prepared five sets of photos, grouped into chapters, exclusively for this purpose. The subjects were also asked to spend five minutes to familiarize themselves with the four layouts and ask any questions.

To eliminate ordering effects from the four layouts, we balanced the user study for each task, *i.e.* the order in which subjects used the four layouts was systematically varied for each task; each subject used a different order from the other subjects for each task<sup>4</sup>. Subjects were also asked to revisit all four layouts with all photo sets when they answer each questionnaire.

## 5. RESULTS

### 5.1 How Do People Organize Their Photos in Each Event?

At the beginning of the user study, we asked subjects to “group the photos into chapters according to their preference and liking”. This allowed us to first observe and later inquire on the criteria they used to decide the chapter groupings. We have gathered three insights into this process:

First, **users value chapter consistency more than the chronological order of the photos**. While past findings have shown that people want their photos displayed in chronological order [25], all but one of the subjects in our study requested that they be allowed to combine non-adjacent chapters in the timeline, effectively displaying the photos out of their chronological order.

Almost all subjects had at least one photo set where in the midst of photos capturing one moment in the event, *e.g.* a performance on stage, there were a handful of photos that did not belong, *e.g.* photos of the audience. Another example is where in the midst of scenic photos of a nearby landscape, there were photos of friends and/or family. In these cases, subjects wanted to keep all but the handful of photos in one chapter. This observation is similar to how

<sup>4</sup>There are 24 distinct permutations in ordering the four layouts.

people keep printed photos in albums in chronological order, but with small adjustments done for aesthetic reasons [24].

By allowing the subjects to create meaningful chapters as the organizational unit for their photos, what becomes important to them is the consistency of the photos within each chapter. In explaining why they wanted certain photos taken out of a chapter, subjects said that the photos “do not belong there”. This importance supercedes displaying the photos in chronological order. Some subjects mentioned that they “don’t really care” if the photos are not in chronological order, that “sometimes [it] is not that important”.

Secondly, **criteria for chapters include moment, object, location, photography type, and intention**. These criteria pertain to the kind of consistency discussed in the first point. From our study, we observed that the subjects commonly adopted one of the following five criteria for their chapters:

1. **Moment** – This criteria is the most common and refers to chapters that correspond to moments in the event. Several subjects refer to photo sets whose chapters followed this criteria as being “according to time”.
2. **Object** – Subjects wished to group photos of the same object or object type in the same chapter. For example, in a photo set of a trip to a defunct railroad, the subject wanted all photos depicting the track in its own chapter, regardless of when the photos were taken.
3. **Location** – Subjects also commonly organized their photos with a chapter for each location, for example, in holiday photos where photos were captured from a variety of different locations (*e.g.* tourist spots).
4. **Photography type** – For example, subjects wished to group photos of their friends in the same chapter. Another example is to have a chapter for all the scenic photos.
5. **Intention** – On several occasions, subjects wished to have a different chapter for photos of different groups of individuals, *e.g.* one chapter for photos with friends and another chapter for photos with colleagues. Another example is where one subject has several “silly shots” taken at very different times during the event but would like to have them all in the same chapter.

Lastly, **choice of criteria and granularity for segmentation are very subjective**. We found that deciding a criteria for the chapters is a very subjective process. For example, in a photo set of performances on stage, the subject *separated visually similar photos* into several chapters to have one chapter for each performance. On the other hand, another subject wanted to *combine visually similar photos* of different speakers into the same chapter to create a summary of the event in a single chapter. Several subjects noted that they would group photos of the same location, even if taken at different times, *e.g.* night and day, into the same chapter. However, they will separate portrait photos of their friends/family into a different chapter, separate from the chapter with scenic photos of the same location.

Subjects also had different notions of granularity for their chapters. One subject wanted to create a chapter with many photos to depict “photos of the path [he] took from the entrance to the mountain”. Photos taken near the path would

Initialized?	Num Photo Sets	Average $Pr_{miss}$	Average $Pr_{fa}$	Average $Pr_{error}$
No	30	0.193	0.508	0.350
Yes	47	0.118	0.290	0.204
Improvement		38.7%	43.0%	41.8%

**Table 1: Comparison between the chapter groupings by our algorithm with the ground truth by the subjects as measured by miss rate,  $Pr_{miss}$ , false alarm rate,  $Pr_{fa}$ , and error rate,  $Pr_{error}$ . A smaller number indicates better agreement. One group of photo sets were initialized by our algorithm and further organized by the subjects. The other was done by the subjects without help.**

be grouped into separate chapters. Another subject mentioned that he would like to group his photos “by visual similarity” unless “[the photo set] is for a big event because there will be too many chapters”. Many subjects disliked having a chapter with just one or two photos and would combine the chapter with an adjacent one simply because s/he “want[s] to combine it with something else”.

While deciding the chapter grouping is a subjective process, subjects agree that “grouping [their] photos by chapter makes sense” ( $\mu=4.3$ ,  $\delta=0.6$ ). In response to the subjectivity, more subjects found it “easy to decide the correct chapter groupings” ( $\mu=3.7$ ,  $\delta=1.0$ ). These subjects said that they will know what to do when they see the photos.

To assess how automatically grouping photos into chapters affected their final organization by subjects, at the beginning of the study we randomly selected two photo sets from each subject for s/he to group without the help of our event photo stream segmentation algorithm. The other two photo sets of each subject were initialized with a chapter organization given by our algorithm. This allows us to compare the chapter groupings from our algorithm with those by the subjects (as ground truth) for two kinds of photo sets: 1) photo sets that were organized by the subjects without help<sup>5</sup>, and 2) photo sets that were initialized by our algorithm and further organized by the subjects.

Some photo sets were from older generation cameras that did not embed photo metadata<sup>6</sup> in the image files. Since the metadata is necessary for our event photo stream segmentation algorithm, we could not run our algorithm on these photo sets. For this initialization analysis, we have a total of 7073 photos in 77 sets.

To perform the comparisons, we used the error rate metric,  $Pr_{error}$ , proposed by Georgescu *et al.* [11]. This metric improves on *WindowDiff*, previously used by Naaman *et al.* [21] to evaluate their photo stream segmentation method. A lower  $Pr_{error}$  indicates better agreement with the ground truth by the subjects; a score of 0 indicates perfect agreement.  $Pr_{error}$  is an average of the miss and false alarm rates. As such, a method that proposes no chapter boundaries or proposes chapter boundaries everywhere will have an error rate of about 0.5.

In our previous work [13], we noted that our event photo

<sup>5</sup>We ran our algorithm on these photo sets but the results were neither used nor shown to the subjects.

<sup>6</sup>Exchangeable Image File Format (EXIF) data; <http://exif.org>

stream segmentation algorithm has a tendency to propose more fine-grained segmentations. We can see this in Table 1 where the false alarm rate,  $Pr_{fa}$ , is markedly smaller – a 43% improvement from a high rate of 0.508 – for the initialized photo sets. With initialization, subjects were provided with the opportunity to explicitly agree or disagree with our fine-grained results. The effect is that subjects found meaningful chapter boundaries among the many proposed. Without initialization, subjects had to find meaningful chapter boundaries for themselves, resulting in higher false alarm rates for our algorithm in comparison.

While the error rate values we report in Table 1 were computed by penalizing misses and false alarms equally, we found through our user study that in practice, having a high miss rate is more detrimental to the user experience than having a high false alarm rate. Many subjects in our study mentioned during the interview that it was easier to decide if two chapters should be combined than to decide how to split up a chapter. To correct a false alarm is a one-step process of combining the two chapters. But to correct a miss, the user must first realize that there is a miss, then figure out the best position to split the chapter.

## 5.2 How Does Chapter-based Photo Organization Affect The Study Tasks?

In this section, we present quantitative and qualitative results from each task of the study. We also present the level of statistical significance of the quantitative results, *i.e.* the p-value from a two-tailed paired student’s t-test in comparison with the plain grid layout. While our findings have different levels of significance, we note that most are significant at  $p<0.005$ . We present the subjects’ mean response values from the questionnaire in Table 2 for easy reference. Values that are statistically significantly in comparison with the plain grid layout are shown with their p-values in subscript. We elaborate on the tabulated results in the following subsections, but defer comparisons between the three chapter-based layouts to Section 5.3.

### 5.2.1 Task 1: Story-telling from familiar event photos

Subjects agree that “having chapters helps present the event’s story for sets with many photos” ( $\mu=4.3$ ,  $p<0.001$ ). We obtained similar results for sets with few photos ( $\mu=3.9$ ,  $p<0.05$ ), but less statistically significant. When asked for each layout specifically however, subjects agree that each of the chapter-based layouts helps present the event’s story for sets with many or few photos, all with  $p<0.005$ .

We also asked the subjects whether having chapters helps them remember what to say about the event. One subject said that the chapters “help give focus” in remembering. Subjects agree that “having chapters helps [them] remember the event’s story” for sets with many or few photos ( $\mu=4.7$ ,  $\mu=4.1$ ; both  $p<0.005$ ). When asked for each layout specifically, subjects agree that each of the chapter-based layouts helps them remember the event’s story for sets with many photos ( $p<0.001$ ). We obtained similar results for sets with few photos, but only the grid-stacking and space-filling layouts are statistically significant at  $p<0.001$ ; the bi-level layout is less statistically significant at  $p<0.05$ .

Chapters can guide users with their storytelling. In the plain grid layout where no chapter information is presented, one subject said that s/he was “scrolling, scrolling, scrolling”

Questionnaire Statement	Bi-Level	Grid-Stacking	Space-Filling	Plain Grid
The layout helps present the event’s story for sets with many photos	4.2 <sub>0.005</sub>	4.2 <sub>0.005</sub>	3.7 <sub>0.005</sub>	2.4
The layout helps present the event’s story for sets with few photos	4.1 <sub>0.005</sub>	4.3 <sub>0.005</sub>	4.1 <sub>0.005</sub>	3.2
The layout helps them remember the event’s story for sets with many photos	4.0 <sub>0.001</sub>	4.3 <sub>0.001</sub>	3.9 <sub>0.001</sub>	2.6
The layout helps them remember the event’s story for sets with few photos	4.0 <sub>0.05</sub>	4.4 <sub>0.001</sub>	4.1 <sub>0.001</sub>	3.2
The layout helps to find a photo in a set with many photos	3.6 <sub>0.01</sub>	4.4 <sub>0.001</sub>	3.7 <sub>0.001</sub>	2.7
The layout helps to find a photo in a set with few photos	3.6	4.4 <sub>0.001</sub>	4.0 <sub>0.001</sub>	3.1
The layout helps to interpret photos of an event with many photos	3.9 <sub>0.005</sub>	4.6 <sub>0.005</sub>	4.0 <sub>0.005</sub>	2.9
The layout helps to interpret photos of an event with few photos	3.7 <sub>0.05</sub>	4.4 <sub>0.001</sub>	3.9 <sub>0.001</sub>	3.1

**Table 2: Mean response values from the subjects to various questionnaire statements for each layout. The values follow a standard 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Values that are statistically significant in comparison with the plain grid layout are shown with their p-values in subscript.**

and did “not know where to stop and say something more”. In contrast, subjects use the chapter information presented in the other chapter-based layouts to pace their story. Subjects would refer to a particular chapter and start a part of their story with, e.g. “this chapter is about...”. Subjects also gesture around chapter outlines with their forefingers or cursors in the space-filling layout to highlight the photos relevant to their stories at the time. One subject however, adopted a purely photo-driven storytelling method [2] where s/he would double-click to maximize the photo and subsequently use the navigation keys on the keyboard to go to the next or previous photos.

On average, the grid-stacking layout is most preferred, followed by the bi-level, space-filling and plain grid layouts. The difference in preference between each of the chapter-based layouts with the plain grid layout is statistically significant ( $p < 0.001$ ).

### 5.2.2 Task 2: Find a given photo from familiar event photos

From the measured completion times, we determined the layout that allowed subjects to complete the task the fastest. On average, the space-filling layout was the fastest (7.0s), followed by the plain grid (7.8s), grid-stacking (11.2s), and bi-level (14.2s) layouts. The difference between the grid-stacking and bi-level layouts ( $p < 0.005$ ); and the plain grid and bi-level layouts ( $p < 0.05$ ) are statistically significant. We note that this ranking aligns closely with how well the layouts make use of screen space, making our results consistent with past findings that propose displaying many thumbnails at once to help users with their visual search tasks [25].

While the plain grid layout ranks second for the fastest completion time, subjects actually preferred the plain grid layout the least for this task. On average, the most preferred layout for this task is the grid-stacking layout, followed by the space-filling, bi-level, and plain grid layouts. The difference in preference between each of the chapter-based layouts with the plain grid layout is statistically significant ( $p < 0.001$ ).

We note that subjects were not informed on how fast they performed with each layout. This was done so that their layout preference for this task was not affected by the completion time rankings. The contrast between the layout preference and the completion time rankings suggests that for the task of finding a photo within a familiar set, where the fastest and slowest times only differ by several seconds, completion time does not play a major role for their preference.

One subject noted that for tasks like this, “they like to find the chapter first”. Subjects agree that “having chapters helps [them] find a photo in a set with many photos” ( $\mu = 4.4$ ,  $p < 0.001$ ). We obtained similar results for sets with few photos ( $\mu = 4.0$ ,  $p < 0.05$ ), but with less statistical significance. Subjects also agree that each of the chapter-based layouts helps them find a photo in a set with many photos ( $p < 0.001$ , except the bi-level layout with  $p < 0.01$ ). For sets with few photos, only the grid-stacking and space-filling layouts are with statistical significance ( $p < 0.001$ ).

While the subjects’ layout preference contradicts with the completion time rankings, the behavior to find chapters first before finding the photo is similar to past findings. The same study we quoted above [25] found that when users want to search for a particular photo, they will first attempt to remember the event at which it was taken. In our case, we observed that subjects use the chapter groupings to skip chapters that they know will not contain the photo, and look deeper into chapters that might. This process is easiest to perform with the grid-stacking layout, which is the most preferred layout for this task.

### 5.2.3 Task 3: Interpreting unfamiliar event photos

Subjects agree that “having chapters helps [them] interpret photos of an event with many photos” ( $\mu = 4.6$ ,  $p < 0.001$ ) as well as those with few photos ( $\mu = 4.0$ ,  $p < 0.001$ ). When asked for each layout specifically, subjects agree that each of the chapter-based layouts helps them interpret photos of an event with many photos ( $p < 0.005$ ). For sets with few photos, only the grid-stacking and space-filling layout are statistically significant at  $p < 0.001$ ; the bi-level layout is less statistically significant at  $p < 0.05$ .

We observed that generally, the subjects fall into two groups, each with a different approach to the task. Subjects in the first group rely on gathering a visual overview of all the photos to interpret the event. They would scroll up and down fairly quickly to gather a general idea of the event. For this group, a layout that displays many thumbnails at once is most preferred and not having chapter information presented in the layout is not a loss. One subject disliked the bi-level layout for this reason: “I can’t grasp what’s happening because it [displays] one chapter at a time”. Subjects would give a very general interpretation of the event and only comment for every other chapter.

Subjects in the second group rely on chapter information to guide them through the event photos. Some would still gather a visual overview from all the event photos, but they

would describe each chapter in chronological order: “Here they went to... and then to...”. With the plain grid layout where no chapter information is presented, these subjects are at a loss and “can’t tell if the photos are apart or together” and the photos were “hard to describe”. In contrast, the layouts with chapter groupings “look very organized”.

There were more subjects in the second group. As such, the most preferred layout for this task is the grid-stacking layout, followed by the bi-level, space-filling, plain grid layouts. The difference in preference between each of the chapter-based layouts with the plain grid layout is statistically significant ( $p < 0.001$ ).

### 5.3 What Layout Aspects are Important for Chapter-based Photo Organization?

Among the chapter-based layouts, the grid-stacking layout was the only layout that outperformed some others with statistical significance; and it does so for each task. For helping to present the event’s story for sets with many photos, subjects agree more with the grid-stacking layout than with the space-filling layout ( $p < 0.01$ ). For helping to find a photo in a set with many or few photos, subjects agree more with the grid-stacking layout than with the bi-level layout ( $p < 0.01$ ). For helping to interpret photos of an event with many or few photos, subjects agree more with the grid-stacking layout than with all the other layouts ( $p < 0.01$ ).

Regarding the methods used by the chapter-based layouts to present the chapters, subjects like the grid-stacking layout ( $\mu = 4.6$ ) statistically significantly more ( $p < 0.005$ ) than the bi-level ( $\mu = 3.9$ ) and space-filling layouts ( $\mu = 3.6$ ). They liked how the layout shows the “chapter groupings each in a separate grid”. In all tasks and in overall ranking, most subjects indicated the grid-stacking layout as their top preference. All this suggests that subjects value the strength of the grid-spacing layout – a clear top-down presentation of the chronological order of the chapters – more than the strengths of the bi-level and space-filling layouts.

The bi-level layout features an overview of all the event photos afforded by the film strip of chapter thumbnails. Subjects like the film strip ( $\mu = 4.4$ ,  $\delta = 0.7$ ) as it shows “the flow of the event”. Subjects also found it is easy to navigate the user interface ( $\mu = 4.2$ ,  $\delta = 0.8$ ). On the other hand, for the statement “I do NOT like the wasted screen space”, subjects only somewhat disagree ( $\mu = 2.7$ ,  $\delta = 1.0$ ). This contrast suggests that while subjects like and appreciate having the overview, they prefer not to waste much screen space imposed by the restricting view hierarchy, even if its easy to navigate.

The space-filling layout maximizes screen space usage; minimal screen space is wasted while still presenting the chapter groupings. A number of subjects do value maximizing screen space usage more than the chronological order of the chapters; five subjects chose this layout as their most preferred layout overall. Most subjects however, prefer the grid-stacking layout. These subjects found the space-filling layout to be “confusing”.

## 6. CONCLUSION

This paper explores chapter-based photo organization and reports results – qualitative and quantitative with statistical significance – that advocates its use for personal digital photo libraries.

We developed a photo browser, *Chaptrs*, and integrated the event photo stream segmentation algorithm from our previous work to explore how people organize their photos in each event. Our algorithm helps users by automatically grouping event photos into smaller groups of chapters. We implemented a baseline plain grid layout and three chapter-based photo organization layouts in *Chaptrs* to explore how chapter-based photo organization affects storytelling, searching and interpretation tasks; and what photo layout aspects are important for such tasks.

Our subjects found chapter-based photo organization to be helpful in all three tasks. Our study also revealed how the subjects employed chapters in these tasks. The grid-stacking layout was preferred the most in all three tasks and the baseline plain grid layout was preferred the least. Among the results, the following are our primary findings from the study:

1. Users value chapter consistency more than the chronological order of the photos in grouping photos into chapters
2. Choice of chapter criteria and granularity for chapter groupings are very subjective
3. Having low misses is more important than having low false alarms for automatic event photo stream segmentation
4. Users value chronological order of the chapters more than maximizing screen space usage in photo layouts

While we discovered that the preference for criteria and granularity of our subjects were very subjective, our study also shows that our algorithm helps subjects discover chapter groupings. In future work, our algorithm can be coupled with photo clustering algorithms based on intra-photo similarity measures to help users achieve their preferred criteria for organization. In the current state, our algorithm has a characteristic of producing smaller chapter groupings that users can use as building blocks to form an organization of their criteria or to build up into a more coarse-grained organization. Additionally, our algorithm currently assumes that photos in the user’s personal digital photo library are not part of any overlapping events. Integrating a photo stream alignment algorithm [29] as a pre-processing step for our algorithm would be beneficial for users with such photos.

With our findings on the key layout aspects, we plan to integrate the film strip overview into the grid-stacking layout in the next design iteration of *Chaptrs*. With this next iteration, we will conduct a longitudinal study that will provide us with long-term usage data to support the findings and conclusions we have made in this paper. In addition, we can couple this with a qualitative study that focuses on exploratory divergent cognitive tasks [28].

## 7. REFERENCES

- [1] P. André, M. L. Wilson, A. Russell, D. A. Smith, A. Owens, and m. schraefel. Continuum: designing timelines for hierarchies, relationships and scale. In *Proc. of ACM Symposium on User Interface Software and Technology*, pages 101–110, 2007.
- [2] M. Balabanović, L. L. Chu, and G. J. Wolff. Storytelling with digital photographs. In *Proc. of*

- ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 564–571, 2000.
- [3] B. B. Bederson. Photomesa: a zoomable image browser using quantum treemaps and bubblemaps. In *Proc. of ACM Symposium on User Interface Software and Technology*, pages 71–80, 2001.
  - [4] C. Chen, M. Oakes, and J. Tait. Browsing personal images using episodic memory (time + location). In *Proc. of European Conference on Information Retrieval*, pages 362–372, 2006.
  - [5] Y.-X. Chen, M. Reiter, and A. Butz. Photomagnets: supporting flexible browsing and searching in photo collections. In *Proc. of International Conference on Multimodal Interfaces and Workshop on Machine Learning for Multimodal Interaction*, pages 25:1–25:8, 2010.
  - [6] P.-Y. Chi and H. Lieberman. Raconteur: from intent to stories. In *Proc. of International Conference on Intelligent User Interfaces*, pages 301–304, 2010.
  - [7] S. J. Cunningham and M. Masoodian. Metadata and organizational structures in personal photograph digital libraries. In *Proc. of International Conference on Asian Digital Libraries*, 2007.
  - [8] S. M. Drucker, C. Wong, A. Roseway, S. Glenner, and S. De Mar. Mediabrowser: reclaiming the shoebox. In *Proc. of International Working Conference on Advanced Visual Interfaces*, pages 433–436, 2004.
  - [9] S. Fertig, E. Freeman, and D. Gelernter. Lifestreams: an alternative to the desktop metaphor. In *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 410–411, 1996.
  - [10] D. Frohlich, A. Kuchinsky, C. Pering, A. Don, and S. Ariss. Requirements for photoware. In *Proc. of ACM conference on Computer Supported Cooperative Work*, pages 166–175, 2002.
  - [11] M. Georgescu, A. Clark, and S. Armstrong. An analysis of quantitative aspects in the evaluation of thematic segmentation algorithms. In *Proc. of SIGdial Workshop on Discourse and Dialogue*, pages 144–151, 2006.
  - [12] A. Girgensohn, F. Shipman, T. Turner, and L. Wilcox. Flexible access to photo libraries via time, place, tags, and visual features. In *Proc. of ACM/IEEE Joint Conference on Digital Libraries*, pages 187–196, 2010.
  - [13] J. P. Gozali, M.-Y. Kan, and H. Sundaram. Hidden Markov Model for event photo stream segmentation. In *Proc. of ICME 2012 Workshop on Human-Focused Communications in the 3D Continuum (HFC3D)*, 2012.
  - [14] J. P. Gozali, M.-Y. Kan, and H. Sundaram. How do people organize their photos in each event and how does it affect storytelling, searching and interpretation tasks? Technical Report TRC4/12, National University of Singapore Department of Computer Science, April 2012.
  - [15] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as essence for photo browsing through personal digital libraries. In *Proc. of ACM/IEEE Joint Conference on Digital Libraries*, pages 326–335, 2002.
  - [16] D. Huynh, S. Drucker, P. Baudisch, and C. Wong. Time quilt: scaling up zoomable photo browsers for large, unstructured photo collections. In *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 1937–1940, 2005.
  - [17] D. Kirk, A. Sellen, C. Rother, and K. Wood. Understanding photowork. In *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 761–770, 2006.
  - [18] A. Kuchinsky, C. Pering, M. L. Creech, D. Freeze, B. Serra, and J. Gwizdka. Fotofile: a consumer multimedia organization and retrieval system. In *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 496–503, 1999.
  - [19] T. Mei, B. Wang, X.-S. Hua, H.-Q. Zhou, and S. Li. Probabilistic multimodality fusion for event based home photo clustering. In *Proc. of IEEE International Conference on Multimedia and Expo*, pages 1757–1760, 2006.
  - [20] T. J. Mills, D. Pye, D. Sinclair, and K. R. Wood. Shoebox: A digital photo management system. Technical Report 2000.10, AT&T Research, 2000.
  - [21] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. In *Proc. of ACM/IEEE Joint Conference on Digital Libraries*, pages 53–62, 2004.
  - [22] D. R. Olsen, Jr. Evaluating user interface systems research. In *Proc. of ACM symposium on User interface software and technology*, pages 251–258, 2007.
  - [23] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: visualizing personal histories. In *Proc. of SIGCHI Conference on Human Factors in Computing Systems*, pages 221–227, 1996.
  - [24] K. Rodden. How do people organize their photographs? In *Proc. of BCS IRSG 21st Annual Colloquium on Information Retrieval Research*, 1999.
  - [25] K. Rodden and K. R. Wood. How do people manage their digital photographs? In *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 409–416, 2003.
  - [26] D.-S. Ryu, W.-K. Chung, and H.-G. Cho. Photoland: a new image layout system using spatio-temporal information in digital photos. In *Proc. of SAC*, pages 1884–1891, 2010.
  - [27] P. Sandhaus and S. Boll. Semantic analysis and retrieval in personal and social photo collections. *Multimedia Tools Appl.*, 51:5–33, 2011.
  - [28] A. M. Webb and A. Kerne. Integrating implicit structure visualization with authoring promotes ideation. In *Proc. of ACM/IEEE joint conference on Digital libraries*, pages 203–212, 2011.
  - [29] J. Yang, J. Luo, J. Yu, and T. S. Huang. Photo stream alignment for collaborative photo collection and sharing in social media. In *Proc. of the 3rd ACM SIGMM international workshop on Social media*, pages 41–46, 2011.