Mediating Photo Collage Authoring

Nicholas Diakopoulos and Irfan Essa
College of Computing, Georgia Institute of Technology
Atlanta, GA 30332-0760 USA
{nad, irfan}@cc.gatech.edu

ABSTRACT

The medium of collage supports the visualization of meaningful event summaries using photographs. It can however be rather tedious to author a collage from a large collection of photographs. In this work we present an approach that supports efficient construction of a collage by assisting the user with an automatic layout procedure that can be controlled at a high level. Our layout method utilizes a pre-designed template which consists of cells for photos and annotations applied to these cells. The layout is then filled by matching the metadata of photos to the annotations in the cells using an optimization algorithm. The user exercises flexibility in the authoring process by (a) maintaining high-level control through the types of constraints applied and (b) leveraging visual emphases supported by the layout algorithm. The user can of course provide fine-grained control of the final collage through direct manipulation. Off-loading the tedium of collage construction to a user controlled yet automated process clears the way for rapidly generating different views of the same album and could also support the increased sharing of digital photos in the form of compact collages.

Categories and Subject Descriptors: H5.2 [Information interfaces and presentation]: User Interfaces. - Interaction Styles.

General Terms: Algorithms, Design, Human Factors.

Keywords: Photo Collage, Painting Interface

INTRODUCTION

Photo collages are an effective and compact method for summarizing and sharing digital photo albums. In addition to showing each picture in the album, the layout of the collage can carry additional information such as creating a path through time, a story, or an event. Ethnographic research indicates that people enjoy creating collages or "milestone" projects to mark a special life event such as a birthday, graduation, or wedding [8]. Digital collages can thus act as novel visualizations of picture albums, which might be desirable as memorabilia or to share with friends who attended an event [8, 1]. The problem with both traditional and digital collages however is that it can be quite tedious and demanding to *author* a collage from a large collection of photographs.

Our solution is to provide support for the user to maintain control over the selection of photos in the collage, but offload the time-consuming process of laying out the photos in space to an optimization algorithm that can be directed, at a high level, by the user. Thus the user maintains agency in the authoring process, but is relieved from the tedium of it. Our approach provides the full range of task migratability insofar as fully manual, fully automatic, and mixed modes of authoring are all supported. We believe that a human in-the-loop approach to semi-automated media authoring represents an important paradigm for authoring content from growing digital corpora.

UIST'05, October 23-27, 2005, Seattle, Washington, USA.

Our layout algorithm incorporates the metadata of the photos in the album (*e.g.*, time, faces, keywords) in providing a dynamic layout and view of that album. Users can control which photos are emphasized in a layout at a high-level by specifying keywords such as people or places. Alternatively, users can paint with metadata to have a finer degree of control over where photos are placed. Finally, the collage can be directly manipulated in order for the user to precisely alter the layout.

A novel aspect of our work here is that we provide a tool for the user to author new content from an existing set of content. Building off of concepts from multimedia databases, a particular piece of metadata is selected by the user for emphasis, thus acting like a query which filters and structures the resulting content in the collage. Running at interactive rendering rates, the dynamic query mechanism in our system provides for a novel way of authoring differing final collages from the same set of content. For the purposes of sharing collages, the ability to rapidly re-purpose and personalize media for different recipients represents a powerful communicational device.

MOTIVATION AND RELATED RESEARCH

The need for simple interfaces and presentation strategies for images is motivated by ethnographic studies looking at photo sharing and organization behavior. Balabanovic notes that a common use of photographs is to share stories about experiences, places, and people [1]. In story-driven sharing the user gathers the appropriate photos and then recounts a story. Narrative collages naturally support story-driven sharing since a collage representation of an album compactly communicates the story in a single view. This aspect of a collage as a communication medium motivates us to consider developing tools to support user authoring of collages.

The algorithmic layout of visual content for either artistic, summarization, or browsing purposes is another area of research which informs us. A primary line of prior work on filling space with images has focused on generating meaningful summaries of videos using key frames. Work such as Video Manga [15] sought to provide a comic book like visual

summary of video frames. Girgensohn [9] extends the work and provides several avenues for accelerating the global optimization strategy, though their approach is fundamentally limited to a fixed set of key-frame configurations and sizes.

There have been a number of other approaches to creating still image collage summaries from video such as that of [6]. Mamber talks more theoretically about mapping thumbnails taken from film into a spatial arrangement [13]. He indicates several types of data that might be mapped together with the thumbnails such as: people, places, or objects, and stylistic effects including camera motion, lighting, distance, point of view, and image clarity. In our work we consider people, objects, location, and time as well as within-frame focal points and look vectors in guiding our layout algorithm.

Fogarty's work [7] is an important effort toward creating aesthetic image collages. Several of the aesthetic factors and heuristics defined there and in [5] are incorporated into our own work. Some drawbacks of the simulated annealing layout algorithm chosen there are that the perturbation rules are difficult to encode and weight, and that it can be slow to converge. Our algorithm subverts these issues by providing more structure to the optimization problem. Also, whereas the focus of Fogarty's work was in utilizing an aesthetic template, our work concentrates on using semantic templates incorporating higher level photo metadata and annotations.

The PhotoMesa system provides several compact space-filling layout algorithms designed to cluster or group photos based on directory structure or similarity of other available metadata(*e.g.*, time, keyword)[3]. In the same vain as that work we strive to create unique and effective alternatives for laying out images using metadata, though our work stresses the user's agency in specifying how the metadata will affect the layout. This bears resemblance to Kang's work on *semantic regions*, which leverages human spatial memory for arranging personal media into user specified rectangular regions based on associated metadata [10]. Finally, whereas PhotoMesa was designed to cope with very large collections of images, our application is explicitly geared toward smaller albums (less than 100 images) since in larger collages it can become difficult to distinguish features in the shrunken photos. Our system could be integrated with PhotoMesa or any other general photo organization tool.

METADATA

Metadata is an important part of our application since the layout of the collages is guided by photo metadata. One source of metatdata is EXIF, which is saved by digital cameras and includes a variety of capture context, such as date and time information. IPTC is another type of metadata which is more semantically meaningful than EXIF and includes entities such as the title, caption, location, and keywords associated with the image. Metadata can also be generated automatically using content analysis algorithms. For instance, we utilize a face detector since faces are an important indicator of focus in photography. Additionally, we provide the user with a simple interface for indicating both a focus area, and a look vector of an image. Both of these pieces of information can be incorporated to provide better automatic cropping and more intelligent and meaningful layouts.

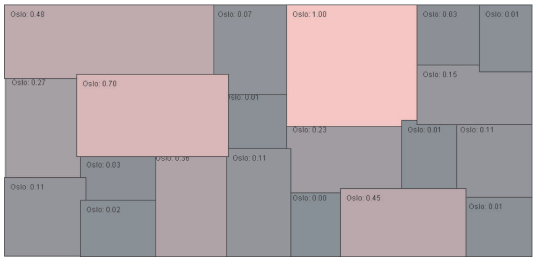

Figure 1: A sample layout template indicating photo cells and associated annotations. The color saturation indicates the weight of the term in that cell.

The content analysis and EXIF metadata is automatic, however the rest of the metadata must be specified by the user. This has the drawback that it can be very tedious to make these annotations. There has been some user interaction research which alleviates this tedium such as bulk annotation [12], or drag-and-drop annotation [14], though it is not the focus of our effort. We assume that the user is willing to provide the necessary additional metadata, whether from another program or though our simple interface. In the future, vision algorithms, GIS databases, or social tagging could provide this metadata at limited expense to the user.

LAYOUT TEMPLATES

To structure the layout algorithm and guarantee that space is filled completely we provide a layout template as input to our algorithm. This template consists of an arrangement of empty cells for each photo to be included. Layout diversity is provided to the user through a library of layout templates offering varying arrangements of cell sizes, shapes, and quantities. Either the user can select an existing template, a suitably sized one can be generated automatically for a given quantity of photos, or the user may design one using a simple WYSIWYG interface. The interface allows the user to directly specify the size, position, and z-order of layout cells. Random overlapping and variation in the depth ordering can be added to give the layout a more ragged look.

Each cell in the layout template typically contains weighted metadata terms indicating the affinity that cell should have toward a photo with the given metadata (Fig. 1). These pieces of metadata are applied as annotations to the cells of the layout template. The user can accomplish this at a fine grained level by using a painting interface, or at a higher level by simply specifying some annotation term that should be emphasized automatically.

Automatic emphasis of an annotation term computes weights for the chosen term applied to all template cells. Users are given the option of emphasizing a term using either the position or size of a cell. Centering a concept is an obvious way of adding visual emphasis, thus for positional emphasis cells near the center are given a higher weight for the term according to a Euclidean distance function. For size emphasis, the area of a cell in relation to the total area of the layout and in relation to the area of the neighboring cells determines the weight of the term for that cell. Incorporating adjacency information, the emphatic power of a large cell placed adjacent to small cells is thus increased and that of a smaller cell placed next to a large cell decreased [5].

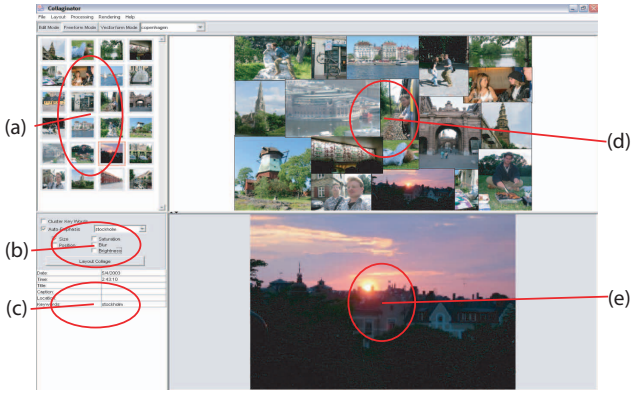

Figure 2: The primary application GUI.

USER INTERACTION

Overview

The application GUI (Fig. 2) is divided into five areas which correspond to the tasks of (a) seeing an overview of photos in the project, (b) changing filtering and rendering options, (c) viewing photo metadata, (d) seeing the overview of the final collage or the annotations on the template, and (e) seeing the selected photo and adding focus or look vector metadata. Dragging a photo from area (a) to area (d) places that photo on the collage. This can be used to construct an entire collage by hand, or simply to anchor important photos and then let the automatic algorithm fill in the rest. To support layout tweaking, drag and drop can also be used to swap individual photos in (d). The collage in (d) acts as a hyperlinked navigator to the rest of the collection; clicking an image on the collage brushes that image in (a), (c), and (e). Area (e) also allows the user to annotate the focus or look vector of an image, and to crop the photo depending on mode. The results of any additional metadata are incorporated in the layout of (d) immediately.

Painting

We utilize the metaphor of painting with keywords for annotating both photos and the layout template. The concept of metadata paint was promoted in the Kandinsky system [7], though painting interfaces have also been explored in earlier work by Baudisch [2]. While in annotation mode, images are annotated by selecting a keyword and then clicking individual thumbnails in (a). The thumbnail border color changes to reflect the addition of the keyword.

Painting over the layout template in area (d) applies the selected keyword to that cell like a can of spray paint, with more paint indicating a higher weight or affinity of that cell for photos with that keyword. Textual feedback indicating the annotation term and its weight, as well as visual feedback indicating weight using color saturation is displayed.

In an alternate annotation mode, the motion of the mouse indicates a path and the paint is applied in a gradient fashion along that path with weights ranging from 0 to 1. This is very useful for specifying time ordering of photos in the space of the collage, since photos will get layed out along the vector according to their timestamp. It can also be used to specify creative positional emphasis different from the automatic central positioning described earlier.

LAYOUT ALGORITHM

Our algorithm serves to optimize the assignment of photos to cells according to the annotations of the photos in the project and the cells in the layout template. A global solution to this discreet optimization problem is far too expensive for our application since we desire near real-time interactive capabilities. We therefore utilize a greedy strategy to initially place photos, followed by a refinement step which selectively perturbs the layout to achieve a more optimal solution.

The optimization of assignment of photos to cells proceeds by minimizing four quantities: a cropping score, σ_c; a focus score, σ_f; a template score, σ_t; and an adjacency score, σ_a. The cropping score measures the amount of a photo which is cropped away by a layout cell. The focus score measures how closely the center of the cropped photo in the layout corresponds to either a focus and/or a face rectangle. The template score indicates how closely the annotation terms of a given photo match the annotation terms of a given cell. The adjacency score measures how well the photos in two cells satisfy any adjacency constraints that have been added. Examples of adjacency constraints that can be specified are “next to”, “looking toward”, and “looking away”. These four scores are summed and optionally weighted by the user to arrive at the ultimate score for each photo mapped to each cell. We make no claim that these are the only or best criteria to optimize, however we did leverage some design principles [5, 7] to guide the development of these score functions. Any other reasonable optimization criteria and score metrics could be implemented and integrated into the layout procedure.

The cropping, focus, and template scores are computed and stored for each photo assigned to each cell at the start of the algorithm. Thus, for a layout with C number of cells and a project with P number of photos there are $C \times P$ initial score computations. The computation of the adjacency score is a bit different since it is dependent on the type of constraints that have been placed on the photos involved. It is still computed between every photo and cell, but the score will change depending on what photos have already been placed in adjacent cells. For the “looking toward” constraint the score function is, $\sigma_a = 1 - \cos\theta$ where θ is the angle between the look vector of the photo and the vector connecting the centers of the cells. For the “next to” constraint the score function is 0 for photos sharing a term which are next to each other and 1 otherwise. Of course all adjacency scores are normalized according to the number of adjacent cells and the number of constraints between the associated photos. To achieve a visual clustering of terms in the layout, a “next to” constraint is added to all photos sharing the same term.

If the user has placed some photos on the layout, these are fixed and taken as the starting point for the automatic process. A sorted list of photo to cell assignment scores is constructed by combining σ_c, σ_f, and σ_t. The photo to cell assignment of lowest score is then used to place that photo in that cell in the layout. The new adjacency scores which incorporate this assignment in the layout are then computed and the list of scores updated and re-sorted. The algorithm proceeds in this fashion until all of the cells have a photo in them.

As the greedy strategy above does not necessarily lead to an optimal solution we adopt a refinement procedure. For each

step of the refinement we consider the global score of the layout (i.e. the sum of all photo to cell scores in the layout). We then compute the $\frac{C\times(C-1)}{2}$ changes in the global score if we were to swap the photos contained in any two cells. The delta score which improves the global score the most is adopted and the corresponding swap is made. This is continued for a set number of iterations or until the score is no longer improving. Satisfaction of the adjacency constraints is greatly enhanced using this process.

RENDERING

There are numerous possibilities for rendering the photos depending on whether a collage or a photomontage aesthetic is desired [4, 11]. For a collage look which clearly delineates the individual photos and their overlaps, we render with borders of various thickness or color. For a smoother photomontage look we specify an alpha blending function which combines the colors of photos in overlapping regions.

Rendering of individual photos within the collage can include any number of image filters as in [7, 11]. Currently we support blurring, saturation, and brightness change for emphasizing keywords so that their associated photos can be visually accentuated and recognized according to the given filter.

RESULTS

Figure 3(a) shows a representative collage generated with the system. It took about 1-2 seconds on a 2GHz Xeon to layout using our algorithm. The pictures have been arranged according to the time vector shown overlaid in (b), which roughly corresponds to the geographic progression of a trip through Ireland. Thus the layout reflects not only time, but gives some cue as to the geographic origin of the pictures. Results showing automatic emphasis, different rendering styles, and the use of look vectors can be found on the project website: http://cpl.cc.gatech.edu/projects/Collage.

In order to get user feedback on the interface and the subjective quality of the collages, we had four users bring in personal photo albums and use the system for about an hour each. They each thought aloud while using the system and then answered a questionnaire. The painting metaphor was found intuitive and unique for controlling the placement of images, though mode switching between painting keywords and direct manipulation of the collage seemed problematic. Two participants expressed a desire for more disorderly layouts with varied image rotations. The saturation, blur, and brightness emphases were helpful for identifying particular images in the collage and the smooth rendering was considered appealing. The look vector constraint was not found particularly useful by any participant. Several participants noted that it was nice to quickly see an initial automatic layout and then tweak it by hand. Finally, among the users who regularly shared photos, such collages were thought of as an easy way of "*sharing a whole album in a single image.*"

FUTURE WORK AND CONCLUSIONS

Evaluation of the usability of our interface and the usefulness of collages as summaries or narratives for sharing is an important area where additional research is needed. One type of study could compare a commercial package to our system and ask users about which solution was seen as more efficient or produced more satisfying results.

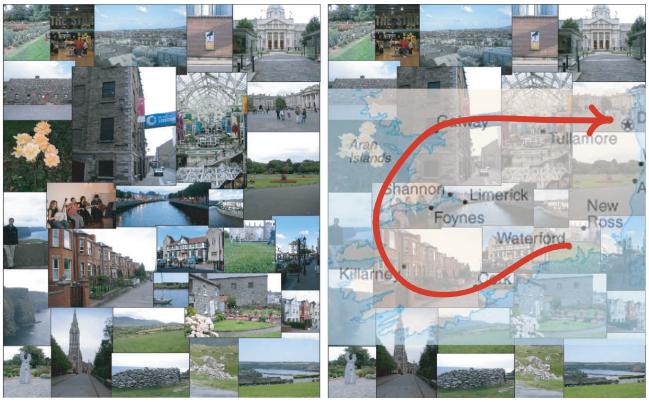

Figure 3: A collage showing a trip through Ireland. See text for explanation.

There are many other applications where the controlled authoring of collages could be applied, such as ambient display of a photo collection, face or object collages, photo essays, semantic contact sheets, or video summaries. We have already begun extending the idea beyond personal photo collections to more general image databases and the creation of "concept collages" for browsing web-based image queries (see project website). Video provides a challenge because selecting meaningful key frames is hard and acquiring the necessary metadata may be even more difficult than for photos.

In summary, we have presented an application and algorithm which empowers the user to create event visualizations and summaries in the medium of collage. As people gather more and more photo data, efficient ways of transforming that data into meaningful and shareable information are necessary. Our effort underscores the need for this and takes steps to address it by giving the user flexibility and efficiency in the collage layout task.

REFERENCES

1. Marko Balabanovic, Lonny Chu, and Gregory Wolff. Storytelling with digital photographs. In *Proceedings of CHI*, pages 564–571, 2000.
2. Patrick Baudisch. Using a painting metaphor to rate large numbers of objects. In *Proceedings of HCI International*, pages 266–270, Mahwah, NJ, USA, 1999.
3. Benjamin B. Bederson. Photomesa: a zoomable image browser using quantum treemaps and bubblemaps. In *UIST '01*, pages 71–80. ACM Press, 2001.
4. George Dillon. Photomontage. http://courses.washington.edu/hypertxt/cgi-bin/12.228.185.206/html/pmontage/.
5. D. Dondis. *A Primer of Visual Literacy*. MIT Press, 1973.
6. Michael G. Christel et. al. Collages as dynamic summaries for news video. In *Proceedings ACM Multimedia*, pages 561–569. ACM Press, 2002.
7. James Fogarty, Jodi Forlizzi, and Scott E. Hudson. Aesthetic information collages: generating decorative displays that contain information. In *UIST '01*, pages 141–150. ACM Press, 2001.
8. David Frohlich and Allan Kuchinsky. Requirements for photoware. In *Proceedings of CSCW*, 2002.
9. Andreas Girgensohn. A fast layout algorithm for visual video summaries. In *Proceedings of ICME*, volume II, pages 77–80, 2003.
10. Hyunmo Kang and Ben Shneiderman. Mediafinder: an interface for dynamic personal media management with semantic regions. In *CHI '03: extended abstracts*, pages 764–765, 2003. ACM Press.
11. Andruid Kerne and Vikram Sundaram. A recombinant information space. In *Proceedings of Computational Semiotics in Games and New Media (COSIGN)*, pages 48–57, 2003.
12. Allan Kuchinsky and Celine Pering. Fotofile: A consumer multimedia organization and retrieval system. In *Proceedings of CHI*, pages 496–503, 1999.
13. Stephen Mamber. Space-time mappings as database browsing tools. *Media Computing: Computational Media Aesthetics*. Kluwer Academic Publishers, 2002.
14. Ben Shneiderman and H. Kang. Direct annotation: A drag-and-drop strategy for labeling photos. In *Proceedings of InfoViz*, pages 88–98. IEEE Computer Society, 2000.
15. S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky. Video manga: Generating semantically meaningful video summaries. In *Proceedings of ACM Multimedia*, pages 383–392, 1999.