



Mixed-Initiative Photo Collage Authoring

Jun Xiao, Xuemei Zhang, Phil Cheatle, Yuli Gao, C. Brian Atkins

HP Laboratories
HPL-2008-199

Keyword(s):

Photo collections, mosaics, image triage, image editing, page layout, constraint satisfaction, user-centered design

Abstract:

Creating an artifact that captures the story or memory from a large photo collection is a difficult task, because the tools available are either too difficult to learn, or oversimplified to the point that they lack flexibility. Individual techniques have been developed to automate parts of the selection-editing-composition cycle, but relatively little has been done to strike the right overall balance between the fully automatic and the fully manual. In this paper, we present miCollage, which attempts to piece together individual technologies to create a compelling collage authoring experience. The system consists of three main components. In the selection component, the system can make proactive suggestions about which photos to add to the collage as well as help the user to find similar or related photos. In the editing component, the system applies automatic cropping and enhancement to the images. In the layout component, the system suggests alternative layouts but is still able to accommodate manual changes, while satisfying various spatial constraints. The user interface connects the components seamlessly, allowing a best of both worlds between fully manual and fully automatic collage authoring.



Mixed-Initiative Photo Collage Authoring

Jun Xiao¹, Xuemei Zhang¹, Phil Cheadle², Yuli Gao¹, C. Brian Atkins¹

¹Hewlett-Packard Labs

1501 Page Mill Road
Palo Alto, CA 94304 USA

²Hewlett-Packard Labs

Filton Road
Bristol BS34 8QZ UK

{jun.xiao2, xuemei.zhang, phil.cheadle, yuli.gao, brian.atkins2}@hp.com

ABSTRACT

Creating an artifact that captures the story or memory from a large photo collection is a difficult task, because the tools available are either too difficult to learn, or oversimplified to the point that they lack flexibility. Individual techniques have been developed to automate parts of the selection-editing-composition cycle, but relatively little has been done to strike the right overall balance between the fully automatic and the fully manual. In this paper, we present miCollage, which attempts to piece together individual technologies to create a compelling collage authoring experience. The system consists of three main components. In the selection component, the system can make proactive suggestions about which photos to add to the collage as well as help the user to find similar or related photos. In the editing component, the system applies automatic cropping and enhancement to the images. In the layout component, the system suggests alternative layouts but is still able to accommodate manual changes, while satisfying various spatial constraints. The user interface connects the components seamlessly, allowing a best of both worlds between fully manual and fully automatic collage authoring.

Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]

General Terms

Algorithms, Design, Human Factors

Keywords

Photo collections, mosaics, image triage, image editing, page layout, constraint satisfaction, user-centered design

1. INTRODUCTION

With the proliferation of digital cameras and memory cards, consumers are taking more photos than ever. However, people rarely consume and repurpose their images beyond individual photo prints. It is not that richer story telling and sharing experiences lack perceived value – people enjoy receiving media creations such as photo collages, calendars and books. Rather, the problem is that for most users, converting a photo collection into

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, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'08, October 26–31, 2008, Vancouver, British Columbia, Canada.

Copyright 2008 ACM 978-1-60558-303-7/08/10...\$5.00.

an artifact that captures the story or memory is difficult, because the tools available are either too complicated to learn, or oversimplified to the point that they lack sufficient flexibility. Consider the example of creating a collage. Most users do not have access to truly flexible photo manipulation and layout software, let alone the time and inclination to develop their own techniques. As a result, in typical solutions, flexibility is traded for the ease of use offered by rigid templates.

In this paper, we present a mixed-initiative photo collage authoring system called miCollage. By “mixed initiative,” we mean that the user retains control over which photos go into the collage, and their positions and sizes; but analysis-based automation is used to alleviate tedious and difficult tasks associated with photo selection, editing and layout. In combination [12], a mixed-initiative system developed by Kerne et al., user and agent collaborate to form and compose image and text surrogates in the course of an information discovery task. Similarly, we believe image analysis based automation should help rather than replace the user, and make the personal photo collage creation task more pleasing, engaging and productive. We seek a middle ground between the fully automatic and the fully manual.

There have been a number of prior works on creating collage summaries from still images. For example, Fogarty defined several aesthetic factors and heuristics and utilized a simulated annealing layout algorithm to obtain the desired result [7]. “Autocollage” creates compact and smooth presentations from photo collections by automatically arranging photos according to Regions-of-Interest (ROI) [17]. Finally, “Tiling Slideshow” displays photos in a sequence of tiled frames synchronized with background music [4]. These systems meet an urgent need for automatic conversion of photo collections into enjoyable artifacts. However, they lack mechanisms of interactivity.

We believe there is great opportunity for designing innovative multimedia applications by considering, from the ground up, interaction models that take advantage of the power and the value of automated image reasoning and at the same time yielding ultimate control back to the user. This semi-automatic approach represents an important differentiation from the previous approach. Our goal is to avoid focusing solely on one or the other, but to seek synergies between the two.

On the one hand, we aim to avoid developing complex analysis and reasoning algorithms only to strip users of their sense of control and accomplishment. For example, which features of the photos are important and of high-value is subjective and may vary greatly between individuals. This is difficult to quantify completely, although image quality heuristics provide some benefit. Furthermore, the photo collection may be dominated by a few major subjects, while a collage calls for a more balanced mix

of characters or scenes. An automatic selection algorithm must be directed by the users in the context of the specific purpose.

On the other hand, we seek to avoid overwhelming the users with decision making and control settings when proactive suggestions can be made automatically. Time consuming tasks such as layout reflow can be well offloaded to an optimization algorithm that maximizes space utility and minimizes disruption. Likewise, some defaults that lead to a sub-optimal yet satisfying presentation can be offered to the user based on fast image analysis results.

Prior work toward human-in-the-loop authoring has been undertaken [5]. However, in this work we focus on the iterative authoring process of image selection, editing and composition. Our review of the literature indicates that while many issues, e.g. automatic cropping, have been explored individually by various researchers, they have yet to be explored in combination with the workflow of the user. Kirk et al. concluded in their “photowork” study that users are mostly likely to carry over extensive selection and editing effort at the time they decide to make use of the photos [13]. However, any editing operation the users made to the individual photos, for example, may very well have a rippling undesired effect on the overall layout of the collage. We endeavor to piece up individual components and attend to every detail of the collage authoring process to create a compelling user experience.

1.1 System Overview

The system consists of three main components. In the selection component, the system can make proactive suggestions about which photos are to be added to the collage as well as help the users to find similar or related photos. In the editing component, the system can apply conservative yet effective auto-crop to the images and enhance their tone and color automatically. In the layout component, the system can provide alternative layout suggestions as well as accommodate changes the users make to the individual photos and the layout, all while satisfying various constraints.

Note that system interactivity requires analysis algorithms that can run in real time on consumer computers. No pre-computed image features were used in our system. In the case of our selection algorithm for example, unlike previous attempts that focus only on precision and recall measures, we aim to provide near instant responses when loading the images. The ability of users to rapidly find, replace, fine tune photos and adjust their layouts on the canvas instantly creates a powerful and flexible experience allowing them to experiment with design options.

One novel aspect of our work is that we provide the automatic solutions in a user interface that seamlessly connects all the components in a natural and intuitive way. Users never have to memorize what the automation tools do or how they work. Most importantly, the user interface creates a fluid transition experience between the fully manual and the fully automatic.

2. FAST IMAGE SIMILARITY AND AUTOMATIC PHOTO SUGGESTION

Among the most tedious and time-consuming tasks in making a collage are sorting the photo collection appropriately, and

selecting photos that best represent the collection. In miCollage, these tasks are alleviated by a design that uses analysis-based mechanisms for finding similar photos, and for recommending photos that best represent the collection.

2.1 Fast Image Similarity

In miCollage, the photo collection is represented in the “photo strip,” a single row of photos along the bottom of the application view, which can be navigated by scrolling horizontally. In many contexts, it is advantageous to order photos according to time stamp [9]; however, in others, such an ordering is either impossible or inadequate. Firstly, metadata – including time stamps – may be absent or inconsistent. For example, it may not have been recorded; it may have been stripped or modified in previous editing; or time stamps from different clocks can disagree. Secondly, users often consider criteria other than time. For example, when assembling a collage from a museum tour, the time dimension may not be as important as the distribution of exhibits visited. In general, sorting photos by content similarity can help users quickly find shots of subjects or scenes, regardless of available metadata. In miCollage, the default order of the photo strip is according to filename, which usually correlates with time; but magic-clicking on a photo in the photo strip will cause the photo strip to be sorted in real time, according to similarity with the magic-clicked photo.

Our focus here was to find a practical similarity algorithm with the right balance of accuracy and speed. This tradeoff is exhibited in existing content similarity metrics, which can be roughly classified according to feature granularity as follows: (a) global features such as color histogram; (b) region-based features extracted from segmented images [2][6][19][21]; and (c) key-point features extracted from interest-point detectors such as SIFT [15]. Generally, finer granularity leads to more accurate results, but at the cost of greater computation. In miCollage, speed was critical, since all analysis would be done at the time photos were loaded into the application. Assuming that the input photo collection would have no more than a few thousand photos, we limited the sorting budget to a few seconds.

Experimenting with well-known algorithms in each of the three categories leads to the following observations. The key-point based similarity algorithms were orders of magnitude slower than what our budget would afford. On the other hand, although the global feature based algorithms were quite fast, they were not sufficiently accurate due to using non-adaptive color quantization. Eventually we settled on region-based algorithms, whose time complexity and accuracy fell somewhat in between and would be able to meet our requirements.

To begin the similarity analysis, we generate segmentations of the images using a fast algorithm to be described in Sec. 3. As a result, each image is represented by an image-dependent set of color clusters. Content similarity can then be measured using the Earth Mover Distance (EMD) [18], which solves for the minimal transportation cost that must be paid to transform one color distribution to the other. Figure 2.1 shows an example sorting result. The algorithm is clearly effective in finding similar as well as near-duplicate photos.



Figure 2.1 Images sorted in order of decreasing similarity to the query, which appears at left.

2.2 Automatic Photo Suggestion

In separate subject studies where we analyzed how users select the photos that best represent a collection, we found that users often employ a two-pass process. In the first pass, they generally get an overview of all the photos, and identify and prune sets of duplicates. In the second pass, they select a “final” set that captures the photo collection and is appropriate for the purpose at hand. Some users can enjoy this step, but many others have observed that it can be time consuming and tedious.

In miCollage, we alleviate two specific photo selection scenarios. In the “auto-populate” scenario, the goal is to automatically generate a complete collage, as a starting point. In the “incremental” scenario, the goal is to select a new photo, from a cluster that is not already represented on the collage if possible.

Formally, both scenarios follow the same process. First, the photo collection is partitioned into clusters of duplicates. This may be followed by a second partitioning, if necessary, to arrive at a set of “suggestion clusters”. Finally, a suggested photo is selected from each suggestion cluster.

In the auto-populate scenario, we first determine the number of photos to appear in the collage. A maximum number of photos $T > 0$ is set beforehand based on the size of the collage, to avoid a crowded result. If the number of duplicate cluster is less than or equal to T , then the suggestion clusters are the duplicate clusters. Otherwise, the suggestion clusters are determined by splitting the sequence at the greatest $T - 1$ similarity gaps.

In the incremental suggestion scenario, the set of duplicates is the set of suggestion clusters. When the user issues a command to add a new suggested photo, each suggestion cluster represented by a photo on the collage is removed from consideration, and a rotating counter is used to identify the next suggestion cluster.

When the suggestion clusters have been determined, each photo is assigned a composite score that is a weighted combination of two metrics described below: typicality within its suggestion cluster, and photo sharpness. For each suggestion cluster, the photo with the highest composite score is deemed the “best representative.”

Duplicate detection: Duplicate detection can be based on similarity only. However, duplicate shots are often taken close in time. In miCollage, time information is used if it is available. We trained two different binary classifiers, each capable of deciding whether two arbitrary photos are duplicates: one using both similarity and time, and the other using only similarity. We measure content similarity using the fast algorithm from Sec. 2.1. Recently the use of key-point based algorithms for near duplicate detection have been heavily investigated and justified [11][22]; however, we did not adopt these approaches due to the very tight computational budget required by our interactive application.

To train the classifiers, we manually labeled 1419 pairs of consumer images, in which 525 pairs are near-duplicates and 894 pairs are non-duplicates. Examples of positive and negative labels are shown in Figure 2.2. Support Vector Machine (SVM) [3] is adopted to train the duplicate detectors with a linear kernel. Ten-

fold cross validation was used to evaluate the accuracy of the resulting detectors.



Figure 2.2. Examples of manual labels for training duplicate detectors. The first row shows positive samples, and the second row shows negative samples.

Using both image similarity and time deltas, we recorded a classification error 12.7% +/- 3.1%, with precision/recall of the duplicate detector 81.1% and 85.7% respectively. Using only image similarity, we recorded a classification error 14.9% +/- 3.8%, with precision/recall of the duplicate detector 81% and 77.9% respectively.

Typicality metric: Intuitively, the most typical image shares the most information with all other images in the cluster. This is equivalent to finding the sample that maximizes its average similarity to the rest of images in the same cluster. In effect, this metric allows us to filter outliers and minimize the propagation of clustering errors to the image suggestion algorithm.

Sharpness metric: Image quality is a general concept that has many dimensions. For example, a good photo should have good exposure, contrast, and color; in addition to good composition, focus on the subject, and pleasing facial expressions. However, given the real-time requirement of our application, our current implementation is mainly focused on avoiding blurry images.

Blur in images often results from motion or lack of focus. Regardless of the cause, blur weakens the major edges in images. For example, in Figure 2.3, the edge strength histograms are shown for two very similar images, one blurry and the other non-blurry. Observe that the edge strength histogram of the blurry image is flatter in shape, and smaller in range than the non-blurry one due to the smoothing effect from out of focus. This observation leads us to the formulation of sharpness score as:

$$Q = \frac{\text{strength}(e)}{\text{entropy}(h)}$$

where $\text{strength}(e)$ is the average edge strength of the top 10% strongest edges and $\text{entropy}(h)$ is the entropy of the normalized edge strength histogram. Intuitively, non-blur images have stronger edges and more peaky edge strength distribution, therefore large $\text{strength}(e)$ and smaller $\text{entropy}(h)$, resulting a larger Q value.

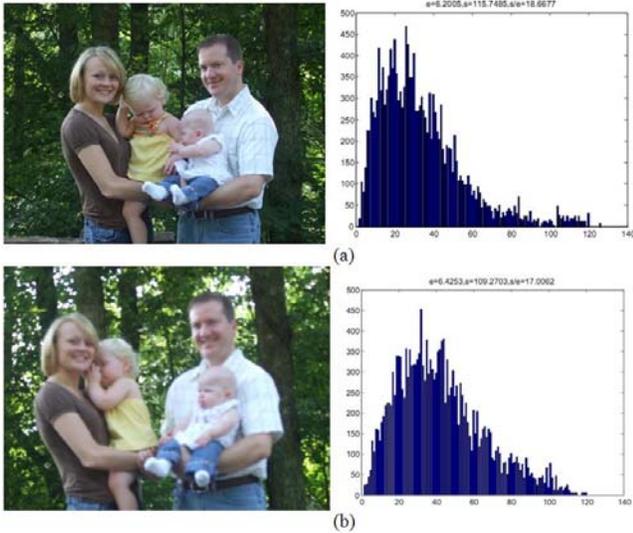


Figure 2.3. Edge strength histograms of non-blurry and blurry images. The non-blurry image has a 18.7 Q score vs. the 17.0 Q score of the blurry image.

3. AUTOMATIC IMAGE ENHANCEMENT

Consumer photographers frequently pay little attention to scene composition. Often the subject is too small, with excess empty space; or distractions at the edges attract the eye away from the main subject area. Appropriate cropping can significantly enhance the visual impact of many images. A by-product of cropping is often a change in aspect ratio which better suits the image content. The miCollage application benefits from this as a variety of aspect ratios produces more interesting collage layouts.

3.1 Autocrop Algorithm

Automatic image cropping consists of two steps: a) image saliency analysis to identify the subject; b) positioning of crop boundaries to include the subject area in an aesthetically pleasing way. Many techniques have been described for automatically identifying salient regions of interest (ROIs) in images. A popular choice is the multi-resolution center-surround difference technique [10]. An original aim of that work was to model the low level attention mechanism in the human visual system. Attempts at extending the technique to ROI location tend to generate rather blurred saliency maps that over-emphasize small edge detail as noted by Liu [14]. We adopt an alternative approach to saliency, based on segmentation. Segmentation produces crisp region boundaries which facilitate the optimization of crop boundary position. We use a segmentation algorithm based on color clustering. This reduces the image to around 20 representative color clusters, Figure 3.1b. Color clusters are assigned a background probability, such that large clusters and clusters close to large clusters (in CIELab) have a high background probability. Each region, (connected component of pixels in the same color cluster), is then classified as “subject”, “background” or “distraction”. A probability for each of these classes is derived heuristically by modifying the region’s color cluster background probability, on the basis of region size and position. Large regions close to the edge are more likely to be background; unusual color, medium sized regions which are relatively central are more likely to be subject; unusual color regions at the edge of the frame are

likely to be distractions. Regions are assigned to the class with the highest probability. This produces a saliency map, Figure 3.1c, where green denotes subject; blue background; and red distraction. A few small regions are unclassified, (grey). Strong region boundaries are also shown in cyan.

We augment this saliency with face detection based on [20] to identify the heads and shoulders of people in the image. The bounding box of all detected people is called the “people box”.

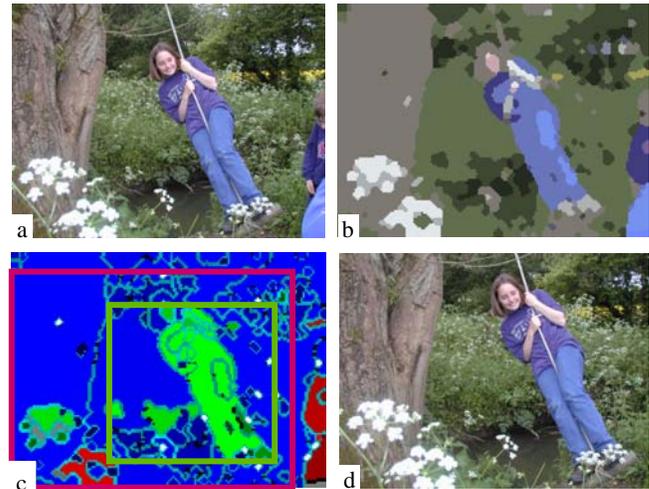


Figure 3.1. AutoCrop: (a) Original; (b) Segmentation; (c) Saliency with minimum crop rectangle (green) and output crop boundary (red); (d) Crop result.

Crop boundary locations are selected by first finding a “minimum crop rectangle” (MCR), shown in green in Figure 3.1c. Then searching the possible rectangle locations which include the MCR using an optimization criterion to select the output crop rectangle, shown in red in Figure 3.1c and hence the crop, Figure 3.1d.

The MCR is created by first forming “subject boxes” - rectangular areas which contain adjacent subject regions. Overlapping subject boxes are merged. Each subject box is scored using the sum of the areas of its subject regions which do not touch the image boundary. The bounding box of these regions is called the “core” of the subject box. The MCR is initially set to the core of the subject box with the highest score. This is expanded to include the people box and the central 15% of the image area. To prevent erroneous cropping of unusual images, the central 25% of the image area is added if the MCR < 20% of the image area, or if the area of subject regions in the MCR < 10% of the image area.

The optimization search finds the crop that minimizes a combination of penalties for: Large crop area; Inclusion of distractions; Proximity to MCR; Proximity to strong region edges parallel to a crop edge; Crossing strong region edges. The penalty function finds crop borders which leave space around the subject, while still producing a reasonably tight crop, rather than simply cropping the ROI as in [16]. For efficiency, a coarse search finds an approximate best crop, followed by a local fine search. Integral images efficiently calculate the penalty criteria during the search.

3.2 Autocrop Evaluation

There are a set of acceptable crops for any image. Solutions get less desirable as more aesthetic conditions are broken. Effective

development and testing requires repeated consistent evaluation over large image test sets. To accommodate this we designed an automatic evaluation method which tests a proposed crop rectangle against crop markup data provided by an expert. Space does not permit a full explanation, but in essence the markup data consists of various types of region such as “must include”, “must exclude”, “boring” etc. The technique has been effective in both developing the algorithm (repeatedly testing over a “training set”), and evaluation over a previously unseen “test set”. Both sets were drawn from the same pool of typical consumer photos. The algorithm can be compared against an “optimal” crop, i.e. the rectangle with the best possible evaluation score, given knowledge of the markup data. Table 3.1 summarizes our results.

Table 3.1. Evaluation metrics over 200 image test set

Evaluation criterion	Optimal	Autocrop
% total pixels cropped	22%	22.6%
% images cropped by >15%	83%	66%
% images unchanged	11%	8%
% images with errors	0%	6.8%

The results are very encouraging. The automatic algorithm makes improvements to most images which require cropping. In some cases it is not sufficiently aggressive. The error cases are typically where part of the “must include” region has been cropped out.

3.3 Automatic lighting/color enhancement

In addition to composition problems, consumer photos frequently have suboptimal exposure and lighting. For most photo creativity applications, color and tone editing is a must-have function.



Figure 3.2. An under-exposed and backlit image (left) after HP RLT image enhancement (right)

In miCollage we provide a one-touch image enhancement option that uses a subset of HP’s proprietary Real Life Technology, which automatically improves images that have contrast and shadow defects. The algorithms bring dark subjects out of the shadows, lighten underexposed images, improve overall contrast, and add saturation to some color regions. Here we used the default setting which has been tuned to improve most images without breaking anything. An example is shown in Figure 3.2.

4. LAYOUT CREATION AND EDITING

In most applications where the user participates in making a photo composite, he or she is provided with a library of static templates. Template libraries are only really appropriate for product types

that contain very few photos. For larger numbers of photos the tasks of producing and using a template library are formidable, since the library should include appropriate ranges of numbers of photos; combinations of aspect ratios; relative photo sizes, etc. Some collage creation solutions address this issue by providing a limited sampling of templates. This often leads to frustrating compromises when the photo set the user wants to show is not supported.

Related work has resulted in automatic composition algorithms that do not rely on predefined templates. In [17] and [8], photos are scaled and positioned so that the canvas is filled completely. These approaches draw irregular boundaries between adjacent photos, making use of content analysis and optimization methods to encourage results in which important regions are not occluded.

Although these methods address the creation of new composites, they are less comprehensive with respect to interactive composite editing. In template based solutions, when photos having different aspect ratios are swapped, they usually are cropped to fit inside their destination regions. This can lead to occlusion of important content. Adding a photo to a page that is already “full” can require that the user replace a photo, or choose a new template. We speculate that the automatic composition methods cited above may present other challenges in this context. For example, replacing one photo with another in which the region of interest is a different shape will have repercussions throughout the design.

We refer to the layout module in miCollage as BRIC, or Blocked Recursive Image Composition. Figure 4.1 presents an overview. BRIC is a method for arranging photos on a rectangular canvas subject to the following primary criteria:

- (a) Photo aspect ratios are respected
- (b) Photo borders and spacing between adjacent photo borders are precisely specified

Criterion (a) reflects the assumption that that process should not impose overlap or other hiding of photographic content unless specifically directed to do so; while criterion (b) provides accommodation for photo borders or other features of graphic design going into the final presentation. Further, creating a new BRIC layout includes the following secondary criteria:

- (c) Photo areas should be proportional to positive relative area values supplied with the photos
- (d) Photos should occupy a maximum of the canvas area

We regard the primary criteria as “musts,” and the secondary criteria as “high wants.”

Referring to the layout in Figure 4.1, note that because of the primary criteria, pairs of adjacent photos and blocks of photos have equal heights or widths. As long as the borders and spacing are not so thick that they take up the entire canvas in either dimension (a weak restriction), any set of photos will fit together in this way. However, we can not guarantee the secondary criteria will be satisfied. Considering criterion (c), for example, notice that the photo with suggested relative area 4 does not have area exactly 4 times that of either photo with suggested relative area 1; and considering criterion (d), although all the photos are large enough to be easily visible, the composite does not completely fill the canvas in the vertical direction.

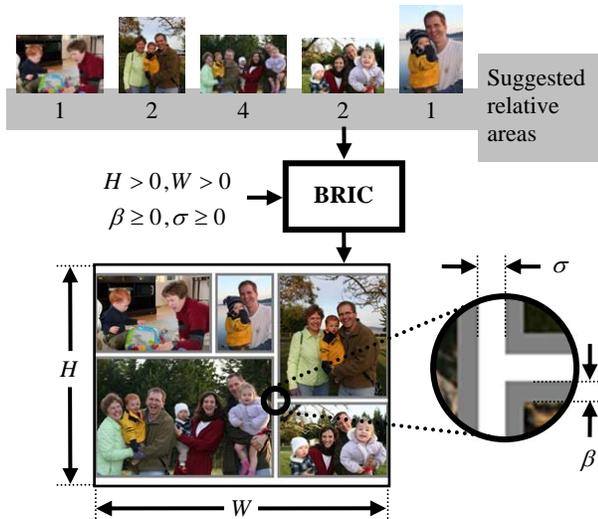


Figure 4.1. BRIC photo composite creation process. In the example, $(H, W, b, s) = (24, 32, 0.25, 0.40)$ inches. The canvas is outlined in black and borders are shown gray.

In BRIC, the composite is encoded as a binary tree which induces a recursive partition of the canvas as illustrated in Figure 4.2. In the tree, each terminal node corresponds to a photo. Each interior node corresponds to a “bounding box” on the canvas, and its designation as either a horizontal (“ H ”) or vertical (“ V ”) cut divides the box into two smaller boxes. Table 4.1 shows the bounding box dimensions for the example of Figure 4.2. Here, for $n \leq 5$, h_n and w_n denote respectively the height and width for a photo; but for $n > 5$, h_n and w_n denote the dimensions of a complete bounding box, including all borders and spacings contained therein. Generally, a table like this can be constructed quickly in a single depth-first traversal of the tree.

Table 4.1. Bounding box dimensions for partition in Fig. 4.2.

Node label	Bounding box height	Bounding box width
$n = 1, \dots, 5$	$h_n + 2\beta$	$w_n + 2\beta$
6	$h_1 + 2\beta$	$w_1 + w_5 + \sigma + 4\beta$
7	$h_6 + h_3 + \sigma + 2\beta$	w_6
8	$h_2 + h_4 + \sigma + 4\beta$	$w_2 + 2\beta$
9	h_8	$w_7 + w_8 + \sigma$

BRIC associates each tree having the form illustrated in Figure 4.2 with a unique layout that satisfies the primary criteria. By *layout* we mean a precise area and position on the canvas for each photo. The essential problem is to compute the photo areas subject to the primary criteria, and this is accomplished by solving a system of N linear equations in N unknowns, where the variables are the photo widths. Referring to the tree, we obtain one equation from each of the $(N - 1)$ interior nodes, by equating dimensions of the bounding boxes of its two children. For example, in Figure 4.2, the constraint furnished by node H_7 would be $w_6 = w_3$. There are two alternative N -th constraints: one equating height of the root bounding box to the height of the

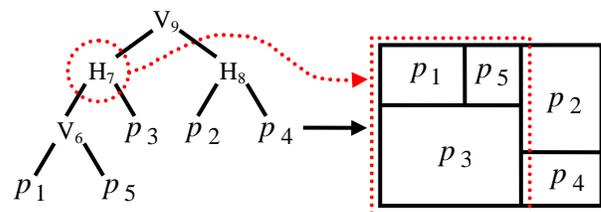


Figure 4.2. Composite is encoded as a binary tree that induces a recursive partition. Interior tree nodes correspond to bounding boxes on the canvas.

canvas, and another for the width. Generally, only one of the two N -th constraints will yield a solution that fits on the canvas.

If we define the aspect ratio of a photo to be its height divided by its width, the coefficients in the linear system are all either $0, \pm 1$, aspect ratios, or negated aspect ratios. As a result, the layout is “continuous”: a small change to the aspect of a photo results in a small change to the layout. For this reason we refer to the process of determining a new layout from a tree structure that has been modified in response to a command from the user as “reflow.”

One important feature of BRIC is that layouts can be created and reflowed very quickly, permitting interactive editing and preview. A summary of the commands supported by the miCollage layout module is given in Table 4.2. A more complete description of the method for creating a new layout is given in [1]. From the foregoing discussion, the table should adequately describe the commands for adding, deleting, replacing, cropping and swapping photos. Figure 4.3 shows an example swap while also illustrating reflow. In the remainder of this section, we sketch out the processes for creating alternate layouts, and for changing the size of a photo.

Table 4.2. Commands supported by the layout engine.

Command	Layout module response
Add or delete a photo	Create a new layout
Replace a photo	Replace terminal node in tree; reflow
Crop a photo in layout	Change aspect ratio of terminal node in tree; reflow
Swap positions of two photos	Swap terminal nodes in tree; reflow
Create an alternate layout	Obtain a new set of photo relative areas; create a new layout
Change size of photo in layout	Described in Sec. 4.2



Figure 4.3. Swap is accomplished by exchanging leaves in the tree, then reflowing the layout.

4.1 Creating alternate layouts

As indicated by Figure 4.1, the inputs to BRIC are an aspect ratio and a suggested relative area for each photo. Our approach to



Figure 4.4 Lower-right photo is continuously grown (to right) and shrunk (to left) by manipulating aspects of other photos.

obtaining a different (or alternate) layout is to submit the photos to BRIC again, but with different suggested relative areas. As described in Section 5, miCollage offers at most three different arrangements. In one arrangement, photos are assigned default relative areas of 1.0, indicating that BRIC should endeavor to find a layout in which all photos have equal area. In a second arrangement, photos with faces are given higher relative areas, and in a third, photos estimated to be of higher overall quality are given higher relative areas. For the sake of simplicity, no check is made to verify that the resulting output is different from the preceding composite. For a vast majority of composites, alternate layouts constructed as described here are different.

4.2 Changing the size of a photo

A photo can be made larger or smaller in the context of a layout by manipulating the aspect ratios of all the other photos, as illustrated in Figure 4.4. To change the size of a photo, the UI identifies the selected photo and supplies a positive side-length factor. Values greater (less) than one indicate growing (shrinking) the selected photo. The side-length factor is applied to the height and width of the selected photo to preserve the aspect ratio.

We first determine target dimensions for the selected photo by multiplying the current height and width by the side-length factor. Changes to the dimensions of the selected photo are translated into target dimensions for the root bounding box. We can now determine new heights and widths for the remaining photos such that target dimensions for both selected photo and root bounding box will be realized upon reflow. For example, in the case of growing, photos separated from the selected photo by a horizontal (vertical) cut will have their aspect ratios reduced (increased).

5. INTERACTION DESIGN

This work was motivated by a desire to investigate the potential of advanced image analysis based automation to address users' needs to offload certain tasks to the computer. It quickly became apparent, however, that there is a need for more research in the multimedia research community that considered user interaction models. For example, automatic image composition techniques are often deployed as tools that simulate human aesthetics and replace manual work in an open loop fashion. However, we would argue that more should be done toward closing the loop enabling two-way interactions between computers and users.

Considering the number of image-analysis-based algorithms that have been incorporated into the system, the user interface could have easily turned into an over crowded toolbox with a dismaying array of features and choices. Such an interface produces a disruptive separation between the photos with which the users are engaged, and the widgets necessary to control them. As a premise our approach was that there should be no menus, no buttons and no toolbars for the users to learn in miCollage. And there should

be no multiple windows or screens to switch back and forth. Every possible action and result should be in plain sight, simple and intuitive. We strove for a minimalist approach to the interface, believing that an excess of visual decoration makes the user acutely aware of the intrusive presence of the computer.

The miCollage design focuses on techniques that enable users to create and refine the collage with fluidity, ease and fun similar to making artifacts such as a physical scrapbook. By deriving visual affordance from physical interactions that occur with printed photographs placed on the tabletop, we intend to minimize the number and complexity of new concepts the users need to learn. Furthermore, we were actively investigating new forms and techniques of input and display beyond the desktop PC. Touch sensitive tabletop displays, for example, seem an ideal context for authoring a photo collage.

All these design challenges led us to the concept of "magic". Magic tricks involve taking something ordinary and making it do something extraordinary. We apply this metaphor by mapping manual to ordinary and automatic to extraordinary. We use "magic" to explain away the complexity of the system and create specific mental models to eliminate perceived complexity. In addition, simple animation tricks make the system seem to run faster than it really does. Our goal is to create a delightful and intriguing user experience by incorporating both whimsy and function. By separating the "magic" from the "ordinary" actions, we expect to see a small mode switching cost from the users. Early user feedback focusing on users' expectations of the action mappings helped us revise the design to minimize the switching cost.

The miCollage application was developed in Flex, allowing it to run across platforms in web browsers and easily access photos provided by web services such as Flickr. The user interface is composed of only four main objects (Figure 5.1). In the upper middle is the magic wand with which the users can apply "magic" operations. Below the wand is the poster board on which users can drag and drop photos to the collage. The bottom left corner has an album icon. To its right is a list of photos in the album.

A minimal set of mouse operations is defined, including left button single click, drag, drop and mouse wheel scrolling. We did not use double click and reserved the right button single click for future undo operations. The effect of mouse operations depends on the context in which it was made, i.e. the object upon which it was made and which mode, magic or ordinary, the user was in.

A list of albums is shown when the application starts. The user can click on an individual album to load its photos. The user can switch between album list view and photo list view by clicking on the album icon. The ordinary and magic modes are toggled by clicking the magic wand icon at the top of the screen. The visual feedback to the user is that the wand is picked up or dropped back. The magic wand is available only after an album finishes

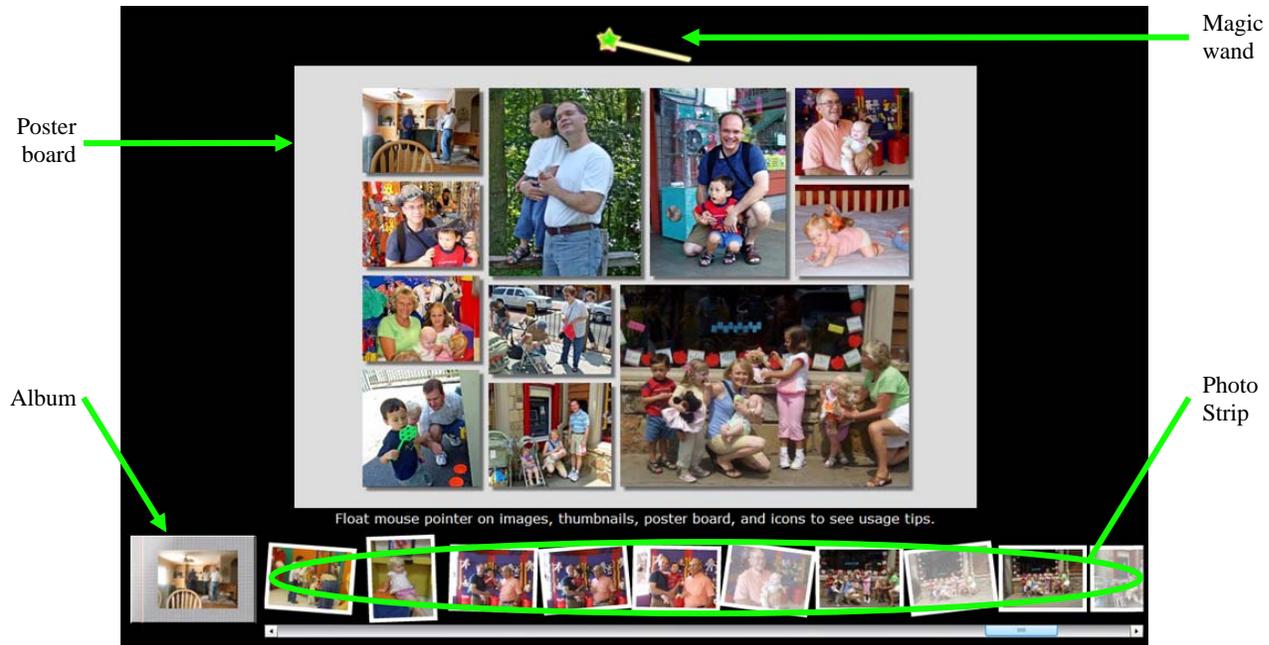


Figure 5.1 The miCollage User Interface

loading. When the mouse pointer hovers over an area, tooltips are shown about the actions that are available on targeted object.

In the ordinary mode, a user can drag a photo from the photo strip to the poster board to add a photo on the collage. A new collage layout is generated immediately. To remove an image from the poster board, the user simply drags and drops the image outside of poster board. Clicking on the poster board background toggles between straight and tilted layout. Scrolling the mouse wheel over an image zooms it in or out, and dragging an image within its frame repositions it relative to its frame. By combining dragging and mouse wheel scrolling operations, the user can manually crop a selected photo on the poster board, although they cannot change the photo aspect ratio yet. Two photos on the same layout can be switched by dragging a photo on the poster board to another poster board image location. The photo may be inappropriately cropped if the target frame has a different aspect ratio.

In the magic mode, on the other hand, when a user drags and drops existing photos on the poster board, the system preserves image aspect ratios and adjusts the layout to accommodate the switch (Fig 4.3). The user can also replace a photo on the board by dragging another photo from the photo strip and dropping on top of the target photo. Clicking on a photo with the magic wand on the poster board toggles auto-crop and auto-enhance on and off for that photo. Scrolling mouse wheel over a photo on the poster board grows or shrinks the size of this photo within the layout, but it may result in cropping of other photos. To add an auto-selected photo to the board, the user can simply click the album icon using the magic wand. Repeated clicks can be used to quickly auto-populate the collage. Clicking the magic wand on a photo on the photo strip sorts all photos by similarity to the targeted photo. Finally, clicking the magic wand on the poster board background cycles through alternate layouts. In our current implementation, we generate 2-3 alternate layouts: one layout uses default target areas, which starts out as equal for all photos but can be changed by user-initiated grow/shrink actions; another layout is generated

using larger target areas for photos with higher image sharpness scores. If there are photos on the collage with faces, a third layout is generated that have larger target areas for images with higher number of faces.

6. EVALUATION

We conducted a user study to evaluate not only the quality of the miCollage user experience but also the effectiveness of the underlying technologies and user interface. Subjects were supplied with a set of photos from the personal collection of someone they know, and asked to create a collage as a gift for the owner using a supplied version of miCollage. We varied the features of photo selection, layout, and editing available to different subjects, and recorded a number of metrics to evaluate user experience and technology effectiveness.

6.1 Experiment design

One goal of the experiment was to collect user action data as they created collages to assess the efficacy of each action. Sixteen kinds of user actions were recorded, such as “add image”, “zoom”, “swap”, “auto-enhance”. These data are less susceptible to demand characteristics of subjective experiments – in our case, tendency of “being nice” by a friendly subject pool.

As a metric for general user experience, we asked subjects to fill out a survey at the end of the collage creation process. They rated overall experience, ease of use, functionality, and satisfaction with the collage layout, each on a 5-point scale. They were also asked to comment on likes, dislikes and missing features.

We also asked the two photo contributors to rank their preferences for each of the collages created by their subjects using a paired comparison procedure. This data provided an “end-result” metric which is unlikely to be influenced by demand characteristics.

To better understand effects of individual technology components, we manipulated 3 independent variables in the experiment:

- (1) Automatic photo suggestion: Half the subjects could only add images manually by dragging a thumbnail to the collage; the other half could also magic-click on the album icon to have an image automatically selected and added.
- (2) Automatic layout: Half the subjects had full BRIC layout functionality; the other half had a template-like layout for each number of photos on the collage, without automatic adjustment of aspect ratio or layout after swap/replace.
- (3) Availability of image enhancement: Half the subjects had access to this function by magic-clicking on an individual image. This enhancement was accompanied by auto-crop when magic-layout functions were available.

Combination of the 3 variables' values resulted in $2 \times 2 \times 2 = 8$ conditions, which were randomly assigned among subjects.

Two colleagues each contributed a personal photo collection for this study. One collection was a weekend family vacation of 161 photos; the other was a family reunion event of 101 photos. Both were direct downloads from the camera without any pre-selection. We used two fixed collections instead of subjects' own photos to minimize confounding effects from variations in collection size and content.

The two photo contributors then invited colleagues who knew them to participate in the study. A total of 16 subjects, 8 for each photo collection, completed the study. They included researchers, managers, and administrative personnel and none of them had prior direct experience with the application.

6.2 Procedures

Subjects were given a link to the web-based user study, and invited to complete the study on their own computers at their convenience.

At the start of the experiment, only the relevant photo album was loaded for each subject. As the photos were loading, subjects were presented with brief on-screen instructions on what functions would be available in the application. Detailed instructions on what happens when clicking/dragging each element in the application were presented as "tool tips" which showed up when subjects floated the mouse pointer over icons, thumbnails, collage area, or images on the screen. Availability of different magic functions depended on the experimental condition a particular subject was assigned to (see section 6.1). A "done" button was visible to the subject when there was at least one image on the collage. When the "done" button was pressed, the final collage was saved, and the survey page presented.

After subject data collection, the two photo contributors were asked to evaluate the collages created for them. Two versions of subject-generated collages were presented at a time, shown side by side in the same layout and format as the subjects created. Contributors chose the version they would prefer as a gift. All possible paired comparisons were done, and the final results were converted to simple rankings by calculating proportion of times a collage was preferred over another in the paired comparison trials.

6.3 Results and analysis

Subjects took between 2.5 and 34 minutes to complete a collage, with an average of 16 minutes. The number of photos on each collage ranged from 5 to 35, with the average being 13.

Subjective ratings from the 4 ratings questions in the survey were generally positive regardless of experimental conditions, typical of survey results from a friendly subject pool. There were no statistically significant differences in ratings as a function of the independent variables. For further analyses, we will look at user action data, collage preference ranking, and survey comments.

Results on auto selection: Four out of 8 subjects with auto-add function used it more than once. Collage preference results were not significantly different between users who used auto-add versus those with matching conditions but didn't use auto-add.

From user action data, 33 (57%) out of 58 auto-added images ended up on the final collage; 148 (69%) of 216 manually added images were kept on the final collage. This indicated good keep rate for auto-selection that is approaching the efficacy of manual-add action (difference in keep rate was not significant).

Of the 2 subjects who specifically commented on auto-selection, one was very positive, and noticed that auto-add skipped duplicates. The other suggested that auto-add load all the selected pictures at once, and then let the user remove unwanted ones.

Results on auto-enhancement: 8 subjects had access to auto-enhancement, 6 used it more than once. Collages produced with and without magic enhancement had the same preference ratings.

Out of 47 auto-enhance actions, 26 (55%) were kept. When auto-enhancement was accompanied by auto-crop, 72% (21 out of 29) were kept. This is significantly higher than the 28% (5 out of 18) keep rate of auto-enhancement without auto-crop ($p < 0.01$, 2-sided), indicating high auto-crop effectiveness.

From survey comments, two subjects specifically mentioned they liked the auto-crop and enhance feature.

Results on magic layout: All 8 subjects who had access to magic layout functions took advantage of at least some of them (swap, shuffle or grow/shrink). The collage preference ratings of collages produced with or without magic layout are significantly different. 6 out of the 8 collages produced with magic layout had higher ratings than their non-magic counterpart, and one had equal rating. The $6.5/8 = 81\%$ preference is significant ($p < 0.04$, 2-sided).

From survey comments, two subjects liked the "ease" and "intuitive" aspect of magic layout. Two subjects commented about aspects of magic layout they didn't like: One said the automatic change of layout when new photo is added messed up a previous layout that he spent a lot of time fine-tuning; another didn't like the way the "grow" function caused other photos to be cropped.

General comments about UI: Users generally commented on the overall application and UI positively. Eight out of 16 subjects mentioned ease of use as something they liked, 4 said it was "intuitive" or "fun". Features subjects listed as "liked" included both manual and "magic" functions, suggesting good user acceptance of mixed initiative design in this application.

There were also plenty of suggestions for improvements. Some were particular to the prototype and easy to fix, such as lack of "undo". Some would need additional research to address, such as

adjustment of layout when a new photo is added without disturbing the existing photo positions too much.

Fully-automatic collage creation: The collages rated by the two photo contributors included an additional collage generated fully automatically. Auto-selection was used to add one image from each cluster (up to 11 clusters). Each image was auto-cropped and auto-enhanced, and the layout generated by BRIC without manual modification. For the 2 sets of 9 collages each, the full-auto versions were ranked the 1st and the 3rd best. This suggests that giving the user an initial fully automatic starting point and then allowing him or her to fine tune it may be a very satisfactory approach.

7. Conclusions

Our goal with miCollage has been to achieve a user experience that makes the best use of image analysis and composition techniques in the context of collage creation.

- Fast image similarity evaluation helps the user find photos similar to a query photo in real time.
- Photo suggestion methods populate a complete collage for a starting point, and suggest a next photo not represented on the collage at any given time.
- Automatic cropping adjusts photo composition to eliminate excess empty space, without setting photo borders that cut across visually important scene elements.
- Automatic composition creates and adjusts layouts on the fly, while adhering to a consistent set of design criteria.

These techniques are woven together seamlessly through design, following the working hypothesis that a mixed-initiative authoring workflow constitutes a better approach for users who care about the quality of the photo collage but lack the skill or time. To test this claim, we conducted user studies on real personal photo collections. User action data indicated very good efficacy and user acceptance of auto-selection and auto-crop functionality. Collage preference ranking data indicated significant advantage of using magic layout functions to improve the subjective quality of the final collage. The real time interactivity afforded by the fast implementation of various algorithms makes miCollage more like a finished product than a research prototype. Open-ended participants' comments for the study were generally positive, especially on the flexibility and the expressiveness of the system, but also pointed to areas that needed improvement. We plan to collect additional user feedback in more realistic scenarios using qualitative observation and interview techniques.

Looking forward, miCollage is a promising step toward an elegant coupling of automated services with user control and manipulation. Design principles applied in miCollage and derived from our findings can be readily applied to other applications such as photo book and slideshow creation. At the same time, automatic techniques may find other avenues toward incorporation into design, such as touch sensitive displays.

8. ACKNOWLEDGMENTS

We wish to thank Darryl Greig for contributing the face detection algorithm and code, Ramin Samadani for contributing the "honest thumbnail" algorithm and code, the Vancouver Imaging Team for contributing the HP RLT image enhancement component, and Dan Tretter for many helpful suggestions.

9. REFERENCES

- [1] Atkins, C. B. Blocked Recursive Image Composition. *ACM Multimedia*, 2008.
- [2] Carson, C., Belongie, S., Greenspan, H., Malik, J. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. on PAMI*, 2002.
- [3] Chang, C.C. and Lin, C.J. LIBSVM : a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [4] Chen, J.C., Chu, W.T., Kuo, J.H, Weng, C.Y. and Wu, J.L. Tiling slideshow. *ACM Multimedia*, 2006.
- [5] Diakopoulos, N., and Essa, I. Mediating photo collage authoring. *UIST*, 2005.
- [6] Fan, J., Gao, Y., Luo, H. Multi-level annotation of natural scenes using dominant image compounds and semantic concepts. *ACM Multimedia*, 2004.
- [7] Fogarty, J., Forlizzi, J. and Hudson, S. Aesthetic information collages: generating decorative displays that contain information. *UIST*, 2001.
- [8] Girgensohn A. and Chiu, P. Stained Glass Photo Collages. *UIST*, 2004.
- [9] Graham, A., H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as essence for photo browsing through personal digital libraries. *JCDL*, 2002.
- [10] Itti, L., Koch, N., and Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on PAMI*, 20(11):1254-1259.
- [11] Ke, Y., Sukthankar, R. and Huston, L. An efficient parts-based near-duplicate and sub-image retrieval system. *ACM Multimedia*, 2004.
- [12] Kerne, A., Koh, E., Dworaczyk, B., Mistrot, J.M., Choi, H., Smith, S., Graeber, R., Caruso, D., Webb, A., Hill, R., Albea, J. combinFormation:a Mixed-Initiative System for Representing Collections as Compositions of Image and Text Surrogates, *JCDL*, 2006.
- [13] Kirk, D., Sellen, A., Rother, C., and Wood, K. Understanding photowork. *CHI*, 2006.
- [14] Liu, T., Sun, J., Zheng, N., Tang, X. and Shum, H. Learning to Detect a Salient Object. *CVPR*, 2007.
- [15] Lowe, D.G., Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.
- [16] Ma, M. and Guo, J. Automatic Image Cropping for Mobile Device with Built-in Camera. *IEEE Consumer Communications and Networking Conference*, 2004.
- [17] Rother, C., Bordeaux, L., Hamadi, Y. and Blake, A. Autocollage. *SIGGRAPH*, 2006.
- [18] Rubner, Y., Tomasi, C. and Guibas, L. J. A Metric for Distributions with Applications to Image Databases. *ICCV*, 1998.
- [19] Smith J.R. and Li, C.S. Image classification and querying using composite region template. *Journal of Computer Vision and Image Understanding*, 75(1-2), 1999.
- [20] Viola, P. and Jones, M. Robust Real-Time Face Detection. *IJCV*, 2004.
- [21] Wang, J. Z., Li, J. and Wiederhold, G. SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries. *IEEE Trans. on PAMI*, 2001.
- [22] Zhang D. Q. and Chang S. F. Detecting image near-duplicate by stochastic attributed relational graph matching with learning. *ACM Multimedia*, 2004.