Phorigami: A Photo Browser Based on Meta-Categorization and Origami Visualization

Shuo-Hsiu Hsu[1], Pierre Cubaud[2], Sylvie Jumpertz[1],

[1] Orange Labs, 4, rue de Clos Courtel BP91226,
35512 Cesson Sévigné Cedex, France
[2] CNAM-CEDRIC, 55 rue du Turbigo, 75003 Paris, France
{shuohsiu.hsu, sylvie.jumpertz}@orange-ftgroup.com, cubaud@cnam.fr

Abstract. Phorigami is a photo browser whose meta-interface visualizes photos by groups according to the analysis of photo contexts. At the core of Phorigami, we proposed a meta-categorization for photo regrouping. This categorization method encompasses the scope of current or expected recognition technologies. Two experiments are conducted by manual classification tasks to study the pertinence of proposed categorization method. We then outline our meta-interface by applying different interaction technique to feature each photo group.

Keywords: Categorization, digital photo collections, content management, interface design.

1 Introduction

For people in this age of digital explosion, building digital personal photo collections becomes indispensable due to the availability of image capturing devices and various free services on the internet. The drastic growth of digital photos either in public shared space or personal collections becomes a "black hole" from the viewpoint of management and has stressed the conventional, WIMP-based, management interfaces. Organizing the collections through categories or themes is an evident way of reducing/filtering the information flux to the user. A posteriori categorization via content analysis techniques is a very promising approach, since the production of metadata is time consuming and inefficient for large public applications. In the work of Rodden, two basic interface features are required to facilitate photo browsing [4]. The first feature is the task-free management technique such as clustering photos by chronological order that add-on camera metadata is utilized. The other feature is the capacity of presenting maximal content by the interface. In [5], low level content-based analysis is applied to present the result of image search which is clustered by similarity; however, photo clustering by similarity may not enhance the efficiency in browsing when users do not have a clear idea about the target in mind. *PhotoTOC* is integrated with time-based and content-based clustering to automatically organize digital photos according to their events [10]. The result of their experiment confirms positive feedback about using an automatic organization technique for the management of digital photo collection. Besides, *FotoFile* [9] and *EasyAlbum* [11] are applied with an annotation and content-based analysis technique to manage people by face recognition. The problem of establishing a relative viewpoint from unordered digital photos is investigated with content-based analysis in [7]. In addition, Jaimes et al. raise the issue of detecting non-identical duplicate photos in consumer photo collections [6]. They propose a content-based analysis algorithm and conduct an experiment by manual classification to observe the performance of proposed algorithm, but no concrete interface is addressed in their study.

Photoware based on content analysis techniques mostly focus on specifying queries and enhancing its performance, but their visualization and interaction techniques still rely on traditionnal, WIMP interface [1]. Some studies such as PhotoMesa [2], Photo tourism [8], Face Bubble [15] and the work of Porta [3] have tried to make up for this shortcoming by integrating unconventional visualization and interaction technique into the photo browser, however appropriate visualization and interaction techniques are still in search. Moreover, the conventional photo management applications for the end user are, to this day, only used to deal with a specific situation of use. The relevance of these specific situations with respect to "everyman" photo collection is rarely addressed in the literature.

2 A Meta-Categorization for everyman photographs

To date, digital cameras provide more automatic "scene-adapted" functions to enhance photo quality and to facilitate shooting photos. For example, the camera interface in Sony Ericsson K800i provides the "shoot mode" and "scenes selection" as illustrated in Fig.1. The shoot mode offers four shooting statuses while six options are configured in scene selection. Apparently, such digital camera interfaces tend to automatically "categorize" digital photos by different contexts. It is to be noted that the metadata of digital photo has been explored to enhance the efficiency of photo retrieval, yet no interface has mentioned the use of this kind of metadata for photo management. We consider that the metadata from those camera modes can be useful for digital photo management and photos taken by different modes from digital cameras can be pre-categorized in the camera metadata. Such automatic categorization can potentially benefit the management of photo collections.

The previously presented camera modes reveal clues for digital photo management and such camera modes can be traced back form user's photo shooting behavior. Therefore we propose a meta-categorization method based on the analysis of the relation between the viewpoint of the photographer and the focused targets. The objective of categorization is to perform an automatic organization of photo contents and to highlight different contexts in the photo collection. This proposed method can potentially be implemented by the state of the art in content-based image analysis algorithm and camera metadata. Therefore, we analyze the photo shooting in terms of two parameters: the movement of the camera and the focused targets in two states (static and mobile). A similar idea of analyzing photos by camera and photographer has been used in Jaimes et al. [6] in terms of camera, scene and parameters of image to detect duplicate digital photos. In our definition, we add a third dimension called "groups" for targets, in order to include the user experience in taking group photos in a reunion. The Table 1 presents the distribution of parameters and associated dimensions.

In Table 1, six classifications are generated as A, B, C, D, E, F and we explain each classification with associated photo scenarios in Table 2. Type A refers to a scenario of a simple static view point where each single photo is focused on a static object from the same environment; for example, a user takes photos of several artworks on an exhibition and the scenario of "multi-view" in type B describes how the photographer changes the position to take a photo of each face of a static object, especially for buildings. Associated examples are illustrated in Fig. 2. Types C and D belong to the motion-capturing scenario where the photographer takes photos of a moving object. The difference between C and D depends on the status of the camera. An example of type C is presented in Fig. 3 where the photographer stays motionless to take photos of walking pedestrians with a bird's-eye view. Type D presents where the photographer moves the camera to trace an object in motion. Likewise, types E and F in Fig. 4 are the group photos in a reunion. It is to be noted that type E refers to photos of a static group while photos concerning the subject replacement [6] or changes in movement are classified as type F. In addition, the double-shot or multi-shot of a digital camera on the same target are what have caused the problems in managing digital photo collections [6]. In our categorization, such double-shot photos in the same context will be possibly categorized in type B, C, D, F while the intention of photographer is merely to take a "better" photo.

Table 1. Movement of camera and focused targets.

Target of photographer \ Camera	Fixed	Mobile
Fixed	**A**	**B**
Mobile	**C**	**D**
Groups(subject replacement)	**E**	**F**

Table 2. Classifications with associated contexts and interaction techniques.

Type	Intention of Photographer	Photo Context	Interaction Technique
A	Simple static view	Panorama	Panoramic Presentation
B	Multi-view		
C	Motion capturing	Action	Animated Photo Presentation
D	Motion capturing		
E	Groups	Social Relation	Simple Folding Presentation
F	Groups in motion and subject replacement		

3 Pilot study and findings

3.1 Objective and procedure

We have conducted two experiments with manual classification to study the relevance of our meta-categorization for personal photo collections. These preliminary results will be used as the reference for implementing the future prototype. The first experiment (E1) aims to discover how many photos can actually be classified from different personal photo collections. Our hypothesis for E1 was that our proposed categorization method could achieve at least a 50% classification rate in each personal photo collection. We conducted a second experiment (E2) to study the variation of personal judgment in the classification result. In E2, ten participants use our method to classify the same photo album. The objective of the second experiment is not only to obtain an average error rate (variation rate) of manual classification but also to examine how subjects execute tasks in terms of personal judgment toward unknown photo albums.

The samples of photo collections are provided by volunteers and the criterion is to collect "everyman" photo album from novices and amateurs. Our sample collection contains twenty personal albums from ten female and ten male photographers, with an age range of 24-34. Participant profiles are university students, junior high school teachers, university administrative staff, engineers, marketing assistant and stewardesses. The size of volunteers' photo collections ranges from 1149 to 14297 digital photos. To lower the size of our collection samples, we choose photos according to the chronological order of their albums to construct a collection with around 1000 items. The final collection size of 20 volunteers ranges from 973 to 1333 photos. The core of our experiments is the manual classification task by participants in which they group photos by highlighting them on printed material according to our proposed method. The procedure of E1 is conducted by the same participant who manually classified 20 personal photo collections. Our participant ran tasks for two collections per day and each photo collection took around 2 hours for classification task. For E2, one photo collection with a size of 973 photos was chosen as our sample. We invited 10 volunteers to manually classify the sample collection. The description of our proposed categorization is presented before each classification task starts. A hard copy of Table 1 is offered to the participant during each task. The volunteers are Masters students from the Universities of Rennes, unlinked with the owners of our sample collection. The duration of the task in E2 is also around 2 hours.

3.2 Results

We counted the number of classified photos by their categories in each experiment and featured the results by pie charts and sequential bars. The pie charts in Fig. 5 present the distribution of each photo type by color and a category G in black is added to present the unclassified photos. The classification rate in E1 is 57% (std. dev. = 7.9%). In E2, the classification rate is 83% (std. dev = 1.3%). These results confirmed our hypothesis that our meta-categorization can classify more than 50% photos in "everyman" photo collection. The result of E1 for the 20 photo albums is illustrated in Fig. 6. Since the subjectivity of participants affects the classification tasks, the rate of E2 is higher than that of E1. Another reason to the high rate in E2 is that two participants thoroughly classified all the photos. For these two experiments, photos are generally unclassified when they are considered as belonging to different contexts. Where photographers may probably only take one photo of a scene, such a situation will result in unclassified photos. We also discovered that some photos belonging to one of the six photo types were not classified due to their wrong chronological order by the original photographer.

To study the variation of participants, we applied the method of matrix permutation [14] to analyze participants' classification flow in sequential bar chart (Fig. 7). The horizontal axis in the bar chart represents the number of photos ranging from 1 to 973 and the vertical axis represents each participant. We selected four overlap areas as illustrated in Fig. 8 and trace back to the associated photo contents. Firstly, the turns of classification flow are related to photo contents; which means that the more different themes the collection has, the more turns it will get. In our sample photo collection, photos in areas A, B, C and D include fewer themes than those in other areas; thus participants can have a clear idea about the photo context. That is why participants have more consistency in classification of photos in these four areas. A contrasted classification result occurs in the area B. Some participants classify photos into category B while the result of most participants' is category D. Such contrast can be explained either as participants' errors or as the variation by the subjective judgment of participants. The sequential bar chart can also be used to study the frequency of the same classification result for each photo. The classification result is visualized in Fig. 7 by seven plot charts according to the photo type. The axis Y of the plot chart represents the number of participants who have the same classification result for the same photo and the axis X represents the number of photos. We obtained a consistent result for the overlap areas of classification results both from the sequential bar chart and the plot chart. We found that the frequency of unclassified photos was more intensive during the first 120 photos where either participants have not yet got used to the classification method or photos are considered unclassified due to various photo themes. Various themes in the photo collection may probably burden the loading in participants' classification tasks. While participants are not the original photographer, they try to reconstruct the context for photos according to their own visual experiences. Participants who have no emotional feedback for the photos can merely categorize photos by visual content analysis and this is basically the same method that a computer algorithm uses for image analysis. Other participants, whose emotional feedback is triggered by photos, tend to establish the photo context more easily. One participant expressed that photos were much easier to classify if he had already been in the scene where the photographer was. In general, the subjectivity effects participants' decisions as to whether or not their emotional cues are triggered by photo contexts. Furthermore, the result also reveals that some participants have difficulties in detecting changes both in the movement of camera and targets; they can only deal with one factor in classification tasks.

4 Origami Visualization

To embody the characteristics of each photo category, we explore appropriate mappings of metaphor for our interaction techniques in terms of two directions: human experience in interacting with paper–like objects and the art of origami. The law of folding in origami art has been applied to solve problems in engineering, industrial design and scientific work [12]. The folding technique of origami art functions in order to represent an object via a minimal presentation. To achieve the maximal content visualization mentioned by Roden [4], the folding mechanism of origami benefits to minimize the space in visualizing a large collection. Similarly, a folding technique is proposed in the work of [16] to space folding for multi-focus interaction. Potentially, the space-saving of the folding technique may be an alternative when visualizing large multimedia contents on small screen devices [17]. In general, we outline our interface, Phorigami, with different origami visualization in Table 2.

For types A and B in Fig. 9, they express a panoramic view of targets; thus the potential interaction technique can be an interactive panoramic photo presentation such as an accordion-like folding technique which compresses the space of panoramic photos into a folding mode and presents an entire panoramic photo by a horizontal photo extension. For motion by frames in types C and D, the potential mechanism is supposed to manipulate motion frames like a GIF animation. The related technique such as RSVP specializes in presenting images with a specific speed; however, such preset browsing speed fails to be adapted in different contexts and may limit the user behavior in browsing images. Therefore, we propose a "manual" Rolodex technique that users can customize the browsing speed to render the original scenario of motion photos. In addition, a pile-like presentation is used to visualize types E and F.

An interface prototype is currently being developed using the Processing environment, in order to simulate the visualization and the interaction techniques. The three types of origami visualization will be implemented with tactile technology to enhance the sense of manipulation toward virtual objects. Fig.10 shows a snapshot of Phorigami visualizing 561 photos (half of the E2 sample collection, as classified by one of the subjects). In this example, the screen space used by Phorigami is more than three times less than that used by a conventional thumbnail interface. In general, photo albums presented by Phorigami display rich visualization results; however, the usability of the different interaction techniques and related metaphor mapping still need to be examined.

5 Conclusion and Future Work

Photos as a medium convey contextual messages and render the original scenes by viewpoints of photographers. Defending the flood of digital photos necessitates a novel meta-interface whose visualization and interaction techniques are capable of dealing with varied photo types and embodying the characteristic of each category.

While conventional interfaces focus on specifying queries for image retrieval, they fail to satisfy users by customizing functions for varied demand of photo works in browsing. Our proposed meta-categorization and meta-interface are a first step for dealing with everyman photo albums. Our future work will focus on conducting deeper user studies on comparing Phorigami with conventional thumbnail image browsers. The issue of serendipitous browsing will also be considered in the next work steps.

Acknowledgments. We would like to thank our participants at University of Rennes1 and Rennes2 for their participation in the experiment and thanks also goes to Chia-yi Shy, Yi-wen Chen, Po-Wei Huang, Hau-chen Wang, Irene Chen, Pieter-Jan Vandormael and other volunteers for the sharing of their digital photo albums.

References

1. Kirk,D., Sellen,A., Rother, C., Wood, K.R. Understanding photowork. ACM CHI’2006, pp. 761-770
2. Bederson, B.B. PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps. ACM UIST’2001, pp. 71-80
3. Porta, M. Browsing large collections of images through unconventional visualization techniques. ACM AVI’2006, pp. 440-444
4. Rodden, K. How Do People Organise Their Photographs?. BCS-IRSG Annual Colloquium on IR Research (1999)
5. Rodden, K., Basalaj, W. Sinclair, D. and Wood, K. R. Does organization by similarity assist image browsing ? ACM CHI’2001, pp. 190-197
6. Jaimes, A., Chang, S. F., Loui, A. Detection of non-identical duplicate consumer photographs. Pacific Rim conference on Multimedia, pp.16-20 (2003)
7. Schaffalitzky, F., Zisserman, A. Multi-view Matching for Unordered Image Sets, or "How Do I Organize My Holiday Snaps?. ECCV’2002, pp. 414-431
8. Snavely, N., Seitz, S. M., Szeliski, R.. Photo tourism: exploring photo collections in 3D. ACM Trans. Graph. 25(3), pp.835-846 (2006)
9. Kuchinsky, A., Pering, C., Creech, M. L., Freeze, D. F., Serra, B., Gwizdka, J. *FotoFile*: A Consumer Multimedia Organization and Retrieval System. ACM CHI’1999, pp. 496-503
10. Platt, J., Czerwinski, M., Field, B. PhotoTOC: Automatic Clustering for Browsing Personal Photographs, Fourth IEEE Pacific Rim Conference on Multimedia. Vol.1, pp.6-10 (2003)
11. Cui, J., Wen, F., Xiao, R., Tian, Y., Tang, X. EasyAlbum: an interactive photo annotation system based on face clustering and re-ranking. ACM CHI’2007, pp.367-376
12. Robert Lang. Idea + Square = Orgami http://www.ted.com/index.php/talks/robert_lang_folds_way_new_origami.html
13. Cooper, K., Bruijn, O., Spence, R., Witkowski, M. A comparison of static and moving presentation modes for image collections. ACM AVI’2006, pp.381-388
14. Bertin, J. La graphique et le traitement graphique de l'information. Flammarion, Paris (1977)
15. Xiao, J., Zhang, T. Face Bubble: photo browsing by faces. ACM AVI’2008, pp.343-346
16. Elmqvist, N., Henry, N., Riche, Y., Fekete, J.D. Melange: space folding for multi-focus interaction. ACM CHI'2008, pp.1333-1342
17. Hsu, S. H., Jumpertz, S., Cubaud, P. A tangible interface for browsing digital photo collections. ACM TEI'2008. pp: 31-32.

Fig. 1. Interface of Shoot Mode(SM) and Scenes Selection(SS) in Sony Ericsson K800i. http://www.sonyericsson.com

Fig. 2. Type A: Simple static view(left), Type B: Multi-view (right)

Fig. 3. Type C: Motion capturing with fixed camera (left)
Type D: Motion capturing with mobile camera (right)

Fig. 4. Type E: Groups (left), Type F: Groups in motion(right)

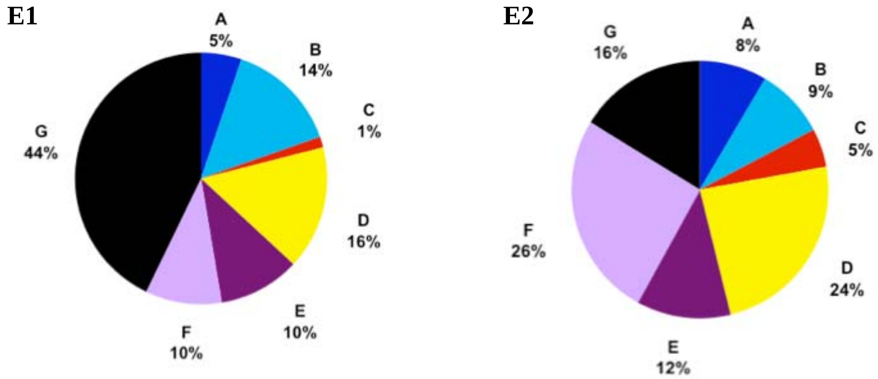

Fig. 5. The distribution of photo types in E1 and E2.

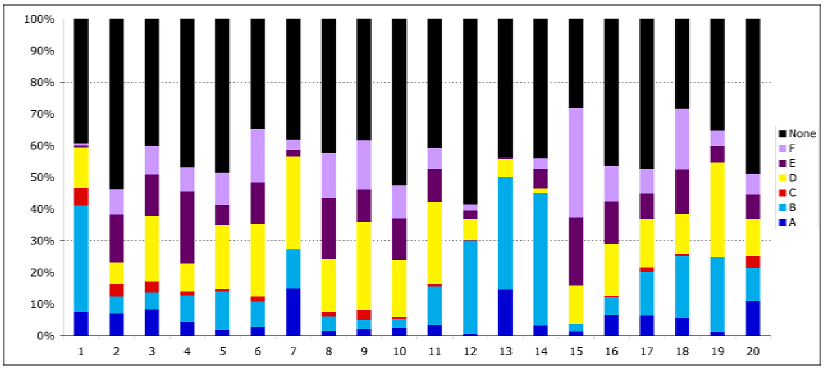

Fig. 6. The result of E1 in detail.

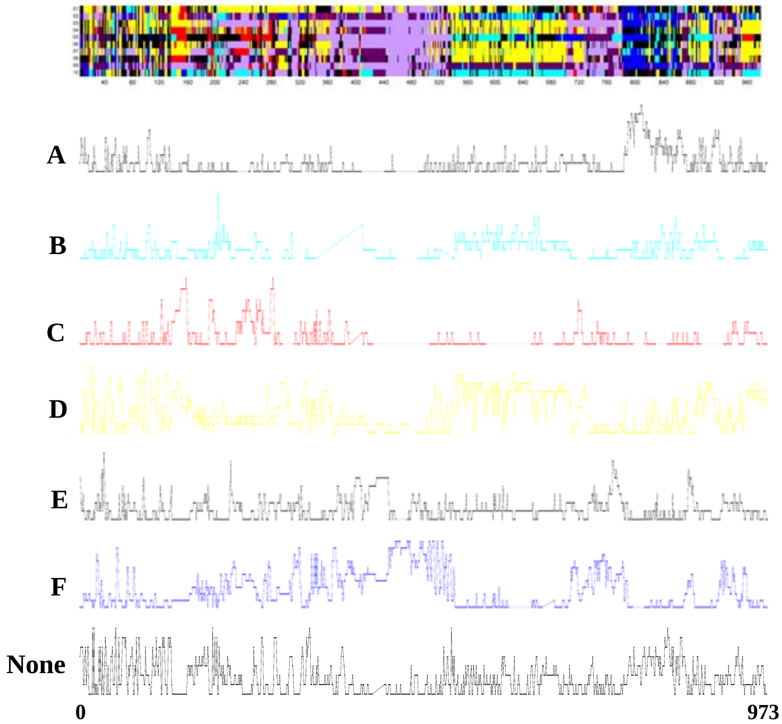

Fig. 7. Sequential bar chart for E2.(top), the color-coding corresponds to Fig.5 Plot chart for E2 (bottom)

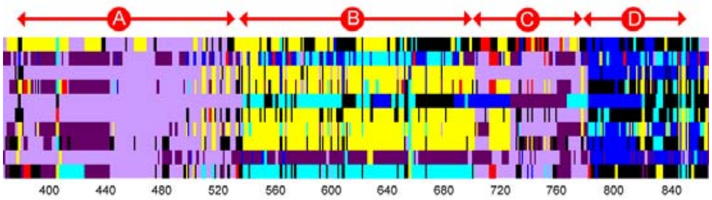

Fig. 8. Overlap areas in the sequential bar chart.

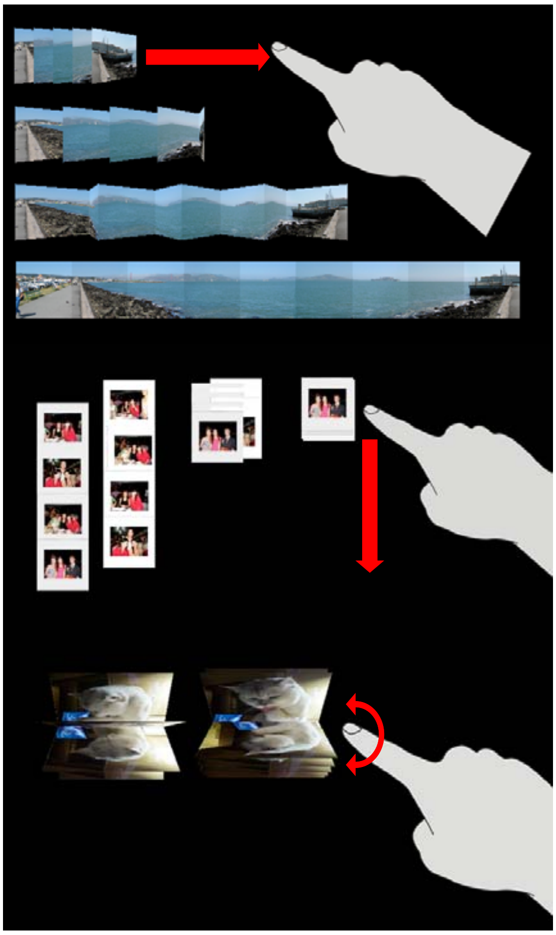

Fig. 9. Panoramic Presentation (top) Group photos (middle) Animated photo presentation (bottom)

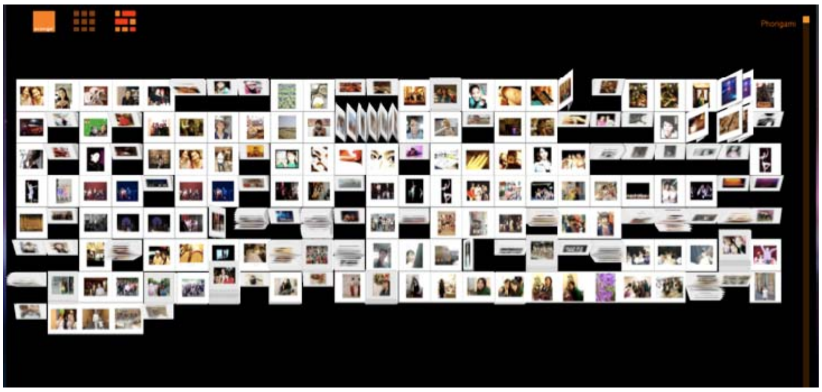

Fig. 10. Visualization of 537 photos with Phorigami.