D-FLIP: Dynamic & Flexible Interactive PhotoShow

Yoshifumi Kitamura[†1] Chi Thanh Vi[†2] Gengdai Liu[†3] Kazuki Takashima[†1] Yuichi Itoh[†4] Sriram Subramanian[†2]
[†1]Tohoku University, Japan [†2]University of Bristol, UK [†3]OLM Digital Inc., Japan [†4]Osaka University, Japan

Figure 1. *Arrangements of photos by metadata.*

1. Abstract

We propose D-FLIP, a novel algorithm that dynamically displays a set of digital photos using different principles for organizing them. A combination of requirements for photo arrangements can be flexibly replaced or added through the interaction and the results are continuously and dynamically displayed. D-FLIP uses combinatorial optimization and emergent computation approach, where parameters such as location, size, and photo angle are considered to be functions of time; dynamically determined by local relationships among adjacent photos at every time instance. Consequently, the global layout of all photos is automatically varied.

Figure 1 (middle) shows an example where all photos belonged to a person are attracted and gathered dynamically around the focused photo when his face is selected. The right figure shows a snapshot when photographs of Mr. A and B were collected in a Venn diagram with red and green circles, respectively. Here, a function of face recognition is assumed to be working and tags for each of the faces are adequately given in advance. Including these, some other examples, which display digital photographs dynamically, interactively, and flexibly, are shown in the video.

2. Background

Pervasiveness of digital cameras has led to large collections of digital photos that users often browse on computer displays, by rearranging them to gather similar ones based on specific features/ meta-data. Although there are several techniques to perform these functions, most of them are somewhat systematic or goal-driven in terms of applying principles for displaying photos. These methods are helpful in organizing and finding photos systematically but previous studies suggest that users often browse their photo collections without a specific search goal [Kirk et al. 2006] but a more general purpose such as looking back at previous experiences or memories. Moreover, users usually browse photos with actions such as enlarging displayed thumbnails in a certain order, displaying photos randomly on a digital photo frame or starting a slideshow for personal gratification and pleasure. To support these behaviors, the presentation of photos should be adapted flexibly and dynamically with visual effects based on user's input.

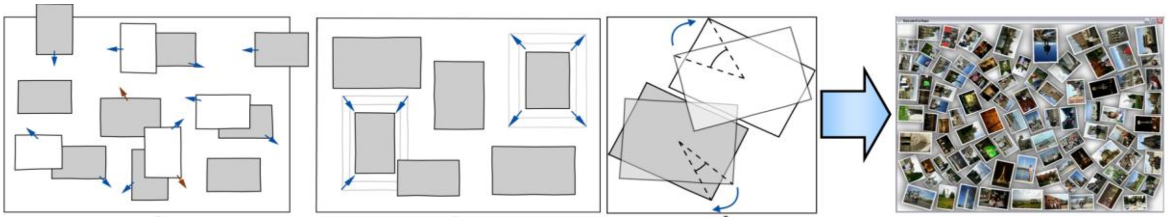
Moreover, many efforts have been proposed in order to arrange photos on a display effectively. For example, a browser that arranges multiple photos in folders by grouping them with different magnification levels [Bederson 2001], by categories with different hierarchy depths [Dachselt et al. 2008], or by displaying more photos in a meaningful way (e.g. adequate groupings [Kustanowitz and Shneiderman 2005]). Other examples are arranging and displaying photos using their shooting dates [Graham et al. 2002], shoot locations [Ahern et al. 2007], shoot locations and people [Gomi and Itoh 2011], and browsing large image datasets using Voronoi diagrams [Brivio et al. 2010]. A technique for browsing large image collections was presented by [Itoh et al. 2004] using the rectangle-packing algorithm, and by [Gomi et al. 2008] using hierarchical tree structured organization of images with level of details (LOD). Additionally, [Chen et al. 2006] proposed an extended slide show in which multiple photos are sequentially displayed in a fixed tiling manner with appropriate music. However, most of these methods allow users to handle photos somewhat in a systematic manner, in terms of selection of requirements or principles of displaying photos. They lack flexibility in displaying with variety of requirements based on user's input.

Additionally, one of the most enjoyable parts of personal photos is to share memories and reminisce with friends or relatives. Previous attempts to provide such experiences have revolved around presenting a static collection with interaction capabilities on tables [Shen et al. 2003] and handheld devices [Balabanović et al. 2000] to facilitate storytelling. However, we want to improve not only the display layout but also the dynamic photo behaviors during interactions. This increases the attractiveness of the representation hence increase users' enjoyment.

Therefore, we propose a novel method to flexibly display a set of photos by showing each of them in a dynamic and continuous motion like a living object. It allows users to replace or add displaying principles interactively and flexibly. As shown in Figure 1, users can manipulate (such as enlarging and translating) a particular photo through flexibly grouping and arranging them using meta-data and/or their feature extracted values. In order to achieve such flexibility, we introduce an approach based on emergent computation. Geometric parameters (i.e. location, size, and photo angle) are considered to be functions of time. Photos are dynamically moved toward the directions determined by local relationships with adjacent photos at each time instance. As a result, the global layout varies automatically and converges gradually with time. At that state, users have an option to freeze the viewing layout or the photos keep moving locally which can be more visually attractive.

3. Proposed Solution

We introduce an approach based on emergent computation to provide the flexibility in photographs viewing methods. In this framework, principles in terms of motions of each photograph can be interactively and flexibly replaced or added. Thus, an interactive and flexible photograph displaying method is achieved. Here, interactions are assumed to be given by input devices such as mice, touch panels, gaze input devices, etc. Photographs are always in motion; with which users enjoy them pleasantly. Dynamic and smooth reactions of photographs by



(a) translation to avoid overlaps and not to exceed boundary + (b) becoming as large as possible and smaller to avoid overlaps + (c) rotation to avoid overlaps ⇒ (d) an arrangement for an interactive digital table surface surronded by multiple users

Figure 2. *Conceptual behaviors of photographs*

interactions provide pleasant and active visual effects that motivate users to see and enjoy photographs.

4. Algorithm Overview

Each photo has three parameters: its position, size, and rotational angle. These parameters are considered as functions of time and are controlled to arrange multiple photos simultaneously on a display. The photo movement is shown by gradually changing the values of these parameters at every time instant. The algorithm is explained by Eq. (1):

$$dx/dt = f(\vec{x}) + \eta \tag{1}$$

Here, $\vec{x}$ is a set of the three parameters above and its variation *dx/dt* is derived by $f(\vec{x})$, the principle to achieve the photo arrangement, and noise term η. Larger amplitude noise increases the fluctuation but is useful for escaping local optimum. In addition, Eq (1) can be re-written with the weight coefficients:

$$\frac{dx}{dt} = \sum_i \{w_i f_i(\vec{x}) + \eta_i\} \tag{2}$$

Here, $f(\vec{x})$, a variety of principles, is used to achieve the photos arrangement or layout. Let P represent the data of a photo, I represents the information of certain input or output devices, $\vec{P}$ is all the photos in the environment, *Position(P)* is the photo position, *Size(P)* is its size, and *Rotation(P)* is its rotational angle. Assuming that the number of principles related to position, size, and rotational angle are l, m, and n. Eq. (3) is obtained from Eq. (2). It controls the parameters of photo P and is calculated from all photos. Here $f_{Pi}(\vec{x})$, $f_{Si}(\vec{x})$, and $f_{Ri}(\vec{x})$ are functions that represent the changes of position, size, and rotation:

$$\begin{cases} \frac{d}{dt} Position(P) = \sum_i^n \{f_{Pi}(I, \vec{P}) + \eta_i\} \\ \frac{d}{dt} Scale(P) = \sum_i^m \{f_{Si}(I, \vec{P}) + \eta_i\} \\ \frac{d}{dt} Rotation(P) = \sum_i^l \{f_{Ri}(I, \vec{P}) + \eta_i\} \end{cases} \tag{3}$$

5. Principles of Photograph Arrangement

There are two types of principles that are important for photo arrangement: *packing* and *mapping*. *Packing* is a geometric problem concerning about arranging multiple photos with different sizes and rotational angles in a pre- determined area; it avoids overlaps and empty regions as much as possible. On the other hand, *mapping* is a semantic concerning about locating each photo based on its content and interaction with users. Here, each function can be established independently based on an individual principle as well as to be implemented without paying attention to the global coordination. Certain feature values of each photo are assumed to be calculated and stored in the tag beforehand (e.g. to specify a person, taken location, etc.). Different arrangements can be achieved flexibly by replacing or adding functions that correspond to the displaying principles.

5.1. Geometric Packing

Here we explain related principles related to geometric packing. First, the principle to avoid overlaps with adjacent photos is represented by Eq. (4). Here, N is the number of photos, *Avoid(P, Pi)* is P's vector for escaping when P and Pi overlap. *Adjacency(P)* is the set of photos overlapping with P.

$$f_{translation}(I, \vec{P}) = \sum_i^N Avoid\ (P, P_i) \quad if\ P_i \in Adjacency(P) \tag{4}$$

Second, a photo is moved toward the inside of the window based on Eq. (5) if its position exceeds the displaying window's border. Here, L, B, R, and T are the left, bottom, right, and top coordinates of the window, *L(P), B(P), R(P),* and *T(P)* are the corresponding photo coordinates, and A_l, A_b, A_r, and A_t are their coefficients, respectively:

$$f_{mold}(I, \vec{P}) = \sum_i^N \begin{Bmatrix} A_l \cdot \{L - L(P_i)\} & if\ L(P_i) < L \\ A_b \cdot \{B - B(P_i)\} & if\ B(P_i) < B \\ A_r \cdot \{R - R(P_i)\} & if\ R(P_i) > R \\ A_t \cdot \{T - T(P_i)\} & if\ T(P_i) > T \end{Bmatrix} \tag{5}$$

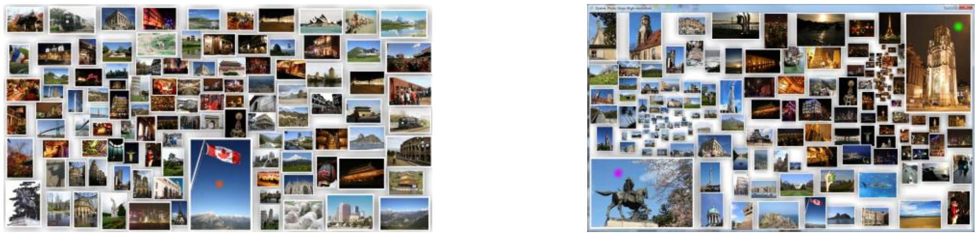
Figure 2 illustrates how photos avoid overlapping. Without overlaps, each photo becomes larger until it reaches the predetermined maximum scale when Eq. (6) is applied (Figure 2a). If two adjacent photos overlap, the larger photo becomes smaller until it reaches the predetermined minimum scale when Eq. (7) is applied (Figure 2b); they move to opposite directions when Eq. (5) is applied (Figure 2a), or rotate in opposite directions when Eq. (8) is applied (Figure 2c). Here, A_{s1} and A_{s2} are coefficients, and $Ang(P_i, P_j)$ is the rotational angle with which P_i and P_j avoid overlapping:

$$f_{enlarge}(I, \vec{P}) = \sum_i^N A_{s2}\{Scale_{\max} - Scale(P_i)\} \tag{6}$$
$$if\ Adjacency(P_i) = \phi \ \wedge\ Scale_{\max} > Scale(P_i)$$

$$f_{shrink}(I, \vec{P}) = \sum_i^N A_{s1}\{Scale_{\min} - Scale(P_i)\} \tag{7}$$
$$for\ \ all\ P_j \in Adjacency(P_i),$$
$$Scale(P_j) < Scale(P_i) \wedge Scale_{\min} < Scale(P_i)$$

$$f_{rotation}(I, \vec{P}) = \sum_i^N Ang(P, P_i) \tag{8}$$

The upper-right photo in Figure 2b will become as large as possible by referring to environmental parameters indicating the positions and sizes of adjacent photos. However, when two photos collide, the larger one becomes smaller, as shown in the lower-right photo based on Eq. (7) if these two equations are simultaneously applied. Thus, all photos are gradually arranged without empty space, and at the same time, their sizes almost become equal if there is not a change of conditions. Even if these two principles conflict, the algorithm will find a solution. Other principles related to geometric packing can be obtained



(a) A photo is focused by overlaying a pointer; (b) Photos are arranged by color and interest similar color photos are collected

Figure 3. *Examples of geometric packing (a) and semantic mapping (b)*

similarly. Figure 2d shows an example arrangement for an interactive digital table surface surrounded by multiple users.

5.2. Semantic Mapping

As one of the simplest examples of semantic mapping, a function to enlarge an interesting photo is represented by Eq. (9). Here, *Attention* is a set of interesting photos given by adequate input device as a mouse or a gaze input device:

$$f_{attention}(I,\vec{P}) = A_{a1} \cdot \{Scale_{\max} - Scale(P)\} \quad if\ P \in Attention \tag{9}$$

Eq. (10) shows how a focused photo attracts other ones with similar attributes. Here, *Similarity*(P_i, P_j) is the similarity between photos P_i and P_j, and if this value is larger than a threshold, P_j moves toward P_i, and away otherwise. The similarities are assumed to be calculated by feature values obtained by image processing or from tags of photos. Related principles of semantic mapping can be obtained similarly.

$$f_{attraction}(I,\vec{P}) = \begin{cases} \sum_i^N A_{a2}\{Similarity(P,P_i) - Threshold\} \\ \quad \cdot\{Position(P_i) - Position(P)\} \\ \qquad if\ Similarity(P,P_i) \geq Threshold \\ \sum_i^N A_{a2}\{Similarity(P,P_i) - Threshold\} \\ \quad /\{Position(P_i) - Position(P)\} \\ \qquad if\ Similarity(P,P_i) < Threshold \end{cases} \tag{10}$$

6. Behaviors of Photographs

Figure 3a shows an example where one photo is focused by overlaying a cursor (at the bottom center of the photo). Figure 3b shows photos arranged by color and user interests using the principles of geometric packing (i.e., Eqs. (4)(5)(6)(7)). In this figure, two cursors (magenta and green) pointing at two photos (bottom-left night scene and upper-right daylight scene). Soon photos with similar colors are moved toward the focused ones. The final layout is achieved by Eq. (10) with semantic mapping principles. Similarly, other feature values calculated by image processing can be used to a group of photos using this principle.

Figure 1 (left) shows an example of photos arranged using Geotags. In this example, photos arranged without overlapping (using the Eqs. (4)(5)(6)(7)) are attracted by their geographical identification metadata (latitude and longitude coordinates) based on Eq. (10) and moved to their corresponding positions on a world map. Figure 1 (middle) shows an example of finding someone's photos. Given a photo set of a social relationship, when a human face in a photo is selected, the size of that photo becomes larger by Eq. (9). Also, all photos containing the selected person are attracted and gathered around the focused photo dynamically by Eq. (10). Here, a face recognition function is assumed to be working and tags for the faces are adequately given in advance. Similarly, Figure 1 (right) displays examples of grouping photos with closed curves drawn by a mouse using meta-data given to each of the photos in advance. In this figure, photos having meta-data of Mr. A and Mr. B are gathered in the red closed curve in the left and the green closed curve in the right, respectively. In the overlapping area of these two closed curves there are photos belonging to both Mr. A and Mr. B.

7. Users' Experience

Our demo provides a new way of viewing digital photographs in which people view dynamically displaying photographs pleasantly, interactively, and flexibly. Experimental results on performance evaluation and user evaluation investigating users' task engagement using EEG in the context of story preparation and telling using D-FLIP is described in [Vi et al. 2013].

References

AHERN, S., NAAMAN, M., NAIR, R., AND YANG, J.H.-I. 2007. World explorer: visualizing aggregate data from unstructured text in geo-referenced collections. *JCDL*, ACM, 1–10.

BALABANOVIĆ, M., CHU, L.L., AND WOLFF, G.J. 2000. Storytelling with digital photographs. *CHI*, ACM, 564–571.

BEDERSON, B.B. 2001. PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps. *UIST*, ACM, 71–80.

BRIVIO, P., TARINI, M., AND CIGNONI, P. 2010. Browsing Large Image Datasets through Voronoi Diagrams. *IEEE Trans Vis Comput Graph 16*, 6, 1261–1270.

CHEN, J.-C., CHU, W.-T., KUO, J.-H., WENG, C.-Y., AND WU, J.-L. 2006. Tiling slideshow. *MM*, ACM, 25–34.

DACHSELT, R., FRISCH, M., AND WEILAND, M. 2008. FacetZoom: a continuous multi-scale widget for navigating hierarchical metadata. *CHI*, ACM, 1353–1356.

GOMI, A. AND ITOH, T. 2011. A personal photograph browser for life log analysis based on location, time, and person. *SAC*, ACM, 1245–1251.

GOMI, A., MIYAZAKI, R., ITOH, T., AND LI, J. 2008. CAT: A Hierarchical Image Browser Using a Rectangle Packing Technique. *Information Visualization*, 82–87.

GRAHAM, A., GARCIA-MOLINA, H., PAEPCKE, A., AND WINOGRAD, T. 2002. Time as essence for photo browsing through personal digital libraries. *JCDL*, ACM, 326–335.

ITOH, T., YAMAGUCHI, Y., IKEHATA, Y., AND KAJINAGA, Y. 2004. Hierarchical Data Visualization Using a Fast Rectangle-Packing Algorithm. *IEEE Trans Vis Comput Graph 10*, 3, 302–313.

KIRK, D., SELLEN, A., ROTHER, C., AND WOOD, K. 2006. Understanding photowork. *CHI*, ACM, 761–770.

KUSTANOWITZ, J. AND SHNEIDERMAN, B. 2005. Meaningful presentations of photo libraries: rationale and applications of bi-level radial quantum layouts. *JCDL*, ACM, 188–196.

SHEN, C., LESH, N., AND VERNIER, F. 2003. Personal digital historian: story sharing around the table. *interactions 10*, 2, 15–22.

VI, C., TAKASHIMA, K., YOKOYAMA, H., LIU, G., ITOH, Y., SUBRAMANIAN, S. AND KITAMURA, Y. 2013. D-FLIP: Dynamic & Flexible Interactive PhotoShow. *10th Advances in Computer Entertainment Technology Conference*.