D-FLIP: Dynamic and Flexible Interactive PhotoShow

Chi Thanh Vi[1], Kazuki Takashima[2], Hitomi Yokoyama[2], Gengdai Liu[3], Yuichi Itoh[4], Sriram Subramanian[1], and Yoshifumi Kitamura[2]

[1] University of Bristol, UK
[2] Tohoku University, Japan
[3] OLM Digital Inc., Japan
[4] Osaka University, Japan

Abstract. We propose D-FLIP, a novel algorithm that dynamically displays a set of digital photos using different principles for organizing them. A variety 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 an approach based on combinatorial optimization and emergent computation, where geometric 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. As a consequence, the global layout of all photos is automatically varied. We first present examples of photograph behaviors that demonstrate the algorithm and then investigate users' task engagement using EEG in the context of story preparation and telling. The result shows that D-FLIP requires less task engagement and mental efforts in order to support storytelling.

Keywords: Dynamic PhotoShow, Emergent Computing, EEG.

1 Introduction

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. While several techniques to do this efficiently exist, most of them are somewhat systematic or goal-driven in terms of applying principles for displaying photos. These methods are useful in systematically organizing and finding photos but previous studies suggest that users often browse their photo collections without a specific search goal (e.g. [1]) but a more general purpose such as looking back at previous memories. Moreover, users often browse photos with actions such as displaying/enlarging photos randomly or starting a slideshow for personal gratification and pleasure. To support these behaviors, the presentation of photos should be flexibly and dynamically adapted with visual effects based on user's input.

Consequently, 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. 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

D. Reidsma, H. Katayose, and A. Nijholt (Eds.): ACE 2013, LNCS 8253, pp. 415–427, 2013.

(a) Photos arranged by Geotags (b) Finding someone's photos (c) Photos corrected by meta-data

Fig. 1. Examples of arrangements by tags

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 of all photos varies automatically and converges gradually with time. These dynamic behaviors provide users enjoyable interactions with less effort to recall good story from the photos. This will enhance one of the most enjoyable parts of personal photos, which is to share memories and reminisce with friends or relatives.

We illustrate example behaviors of photos and then do a user study to evaluate D-FLIP against Windows Explorer, a photo managing program familiar to Windows users. The evaluation involved two participants, a narrator and a listener to prepare and share a story. We measured both participants EEG to quantitatively measure users' metal effort/ task engagement. In addition, NASA-TLX forms were also collected from the narrators and listeners after each task.

The contributions of this paper are: (1) a proposed method to dynamically and flexibly display photos; and (2) an evaluation method using EEG which can be used to evaluate interactive applications.

2 Related Work

2.1 Browsing Digital Photos and Photo Collages

Many efforts were proposed to arrange photos effectively. For example, a browser that arranges multiple photos in folders by grouping them with different magnification levels [2], or by categories with different hierarchy depths [3]. Other examples are arranging photos calendar by using their shooting dates [4], displaying them on a digital geographical map at their shoot locations using meta-data [5], grouping photos with shoot locations and persons [6], and browsing large image datasets using Voronoi diagrams [7]. A technique for browsing large image collections was presented by [8] using the rectangle-packing algorithm, and by [9] using hierarchical tree structured organization of images with level of details. However, most of these methods lack flexibility in displaying with mixtures of requirements based on user's input.

Digital photo collages, which summarize meaningful events or memorabilia, are widely used to display photos. This is efficient because users can view multiple photos at once. However, it requires two types of treatment: (1) a geometric treatment concerning about arranging multiple photos in a pre-determined area but avoids overlapping and empty regions, and (2) a semantic treatment concerning about content of the photos. Several authoring tools have been proposed to create photo collages easily (i.e. AutoCollage [10], Picture Collage [11], and Digital Tapestry [12]).

2.2 The Effect of Animation on Users' Interest

Compared to static method, dynamic photo displaying seems to be more interesting and aesthetically appealing [13]. Previous studies have shown that animation can boost users' performance in learning and teaching such as understanding Newton's law of motion [14]. In other words, animations can help users perform the task (e.g. learning) easier and with a better performance in learning and teaching. In terms of users' interest, animation are likely to increase emotional interest (created by events that are arousing) while static graphics are likely to trigger more cognitive interest (related to the connections between incoming information and background understanding [15]). As the result, D-FLIP may trigger emotional interest from users because of its dynamical and interactive movements. This will help to achieve the goal of D-FLIP which is letting users viewing photos interactively with ease and interest.

2.3 Evaluating Using Neural Signals

Traditionally interactive programs are evaluated by investigating performance or users' behaviors. However, with programs designed for using with ease and pleasure, an evaluation method measuring users' affective or inner states is preferred. Although this can be done by questionnaires answered by participants, they occur after the event when important issues may be forgotten. Neural signals, measured from the brain can better reflect a users' current state and provide an evaluation metric.

There are many methods to detect neural signals such as fMRI, MEG, fNIRS and EEG. A brief summary of those techniques are discussed in [16]. In addition, EEG devices are portable and have high temporal resolution. EEG signals have been also shown to capture the affective state (such as arousal [17] and task engagement [18]).

As one purpose of D-FLIP is to help user browse photos with ease and interest, measuring task engagement can help to evaluate. [19] defined task engagement as the *effortful concentration and striving towards task goals* where task demands and personal characteristics may influence this pattern of processing. Previous studies have shown a positive correlation between EEG engagement and task demands including stimulus complexity processing and the requirement for attentional resources allocation [20]. Consequently, if an application requires low level of task engagement in using, can be considered easier to use when compared with applications requiring higher task engagement. Moreover, [21] present a measure of task engagement from EEG as $\beta / (\alpha + \theta)$. Given this evidence of measuring users' task engagement/ workload, we used it to evaluate D-FLIP by comparing with a competitive program.

3 Algorithm of Dynamic Display of Photos

3.1 Algorithm Overview

Each photo has three parameters: its position, size, and rotational angle. They 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):

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

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 and is useful for escaping local optimum. Furthermore, Eq (1) can be re-written in another form with the weight coefficients:

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

In here, $f(\vec{x})$, a variety of principles, is used to achieve the photos arrangement or layout. Let *P* represents 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*, respectively. Eq. (3) is obtained by modifying 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, respectively:

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

3.2 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 photo arrangements can be achieved flexibly by replacing or adding functions that correspond to the displaying principles.

Geometric Packing: Here we explain 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*.

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

Second, a photo moves 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:

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

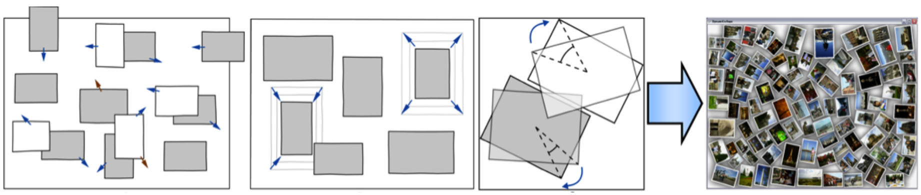
Fig. 2 illustrates how photos avoid overlapping. Without overlaps, each photo becomes larger until it reaches the predetermined maximum scale when Eq. (6) is applied (Fig. 2a). If two adjacent photos overlap, the larger photo becomes smaller until it reaches the predetermined minimum scale when Eq. (7) is applied (Fig. 2b); they move to opposite directions when Eq. (4) is applied (Fig. 2a), or rotate in opposite directions when Eq. (8) is applied (Fig. 2c). Here, A_{s1} and A_{s2} are coefficients, and *Ang(Pi, Pj)* is the rotational angle with which *Pi* and *Pj* avoid overlapping:

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

$$\begin{aligned} &f_{shrink}(I,\vec{P})=\sum_{i}^{N}A_{s1}\{Scale_{\min}-Scale(P_i)\} \\ &\quad for\ \ all\ \ P_j\in Adjacency(P_i), \\ &Scale(P_j)<Scale(P_i)\wedge Scale_{\min}<Scale(P_i) \end{aligned} \tag{7}$$

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

The upper-right photo in Fig. 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 (shown in the lower-left corner) based on Eq. (7) if these two equations are simultaneously applied. Thus, all photos are gradually arranged without empty space while their sizes become almost equal. Even if these two principles conflict, the algorithm will find a solution. Other principles related to geometric packing can be obtained similarly.



(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 at a time instant

Fig. 2. Conceptual behaviors of photographs

Semantic Mapping: 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 an input device as a mouse or a gaze input device:

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

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

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

Viewer's Interactions: Even after the system reaches the balanced condition, the photograph behaviors can be observed when parameters of the display environment vary (e.g. when new photos are added or the size of the displaying window is changed). Also if a cursor (operated by a mouse, for example) is used, the photo overlaid by the cursor becomes larger using Eq. (8) with certain weight coefficients.

Users can observe the displayed photos and interact simultaneously with separate input devices (i.e. touch or gaze input devices). In addition, the display resolution, the number of displays, their positions, orientations, and sizes are variables of the output. Such information about input and output is treated as I in previous equations.

4 Behaviors of Photographs and Performance

4.1 Behaviors of Photographs

The experimental system was developed in C#, using Windows User API and Microsoft .NET Framework 3.5. Some photograph behaviors using the principles explained in the previous section are illustrated here as well as to show the flexibility of our method. They are classified by types: geometric packing and semantic mapping.

a) b) c)

Fig. 3. Geometric packing: a) original layout; b) final layout without rotation, c) with rotation

Fig. 4. Sequence when size of the displaying window is changed

Photograph Behaviors with Geometric Packing: Fig. 3 (a, b) shows how photos avoid overlaps when Eqs (4)(5)(6)(7) are applied. Fig. 3a shows the initial state where 75 photos are located randomly and overlapped. However, they gradually move to avoid overlaps and occupy the empty regions (using Eqs. (4)(5)), as shown from (a) to (b). At the same time, the photos' sizes are varied (using Eqs. (6)(7)), then soon become almost equal (Fig. 3b). Fig. 3 (a, c) is an example of photos avoiding overlaps by rotating in addition to the equations used in Fig. 3 (a, b). In Fig. 3a is the original layout, and Fig. 3c is the layout with collision-free arrangement with rotation. This is useful when photos are shown on tabletop surface displays shared by several users.

Fig. 4 shows an example of photograph behaviors when one of the environmental parameters, the window size, is changed. Once the window is enlarged (left figure), the contained photos steadily move to the empty space (middle figure), according to Eq. (5). This is gradually convergent with time so that the sizes of all photos become almost equal but avoid overlapping (right figure).

Photograph Behaviors with Semantic Mapping: Fig. 5a shows an example where one photo is focused by overlaying a cursor (at the bottom center of the photo). Fig. 5b shows photos arranged by color and user interests using the principles of geometric packing (i.e., Eqs. (4)(5)(6)(7)). Here, two cursors (magenta and green) point 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. (9) with semantic mapping principles. Similarly, other feature values calculated by image processing can be used to a group of photos by applying this principle.

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

Fig. 5. Examples of geometric packing (a) and semantic mapping (b)

Fig. 1a 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. (9) and moved to their corresponding positions on a world map.

Fig. 1b 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. (8). Also, all photos containing the selected person are attracted and gathered around the focused photo dynamically by Eq. (9). Here, a face recognition function is assumed to be working and tags for the faces are adequately given in advance. Similarly, Fig. 1c 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 closed curves there are photos belonging to both Mr. A and Mr. B.

4.2 Discussions

Our proposed method has each photo arrangement generated at every time instant which may not always optimally satisfy all principles. However, these arrangements provide viewers a dynamic photo-viewing environment where they can observe the smooth transitions of photos layout and the behaviors caused by their interactions.

Theoretically, the displayed photos by our algorithm are slightly vibrated because the function inherently includes a noise term in Eq. (1) that causes the system to be constantly in motion. In the above examples, a fixed small amplitude was used for the noise. Larger amplitude noise increases the fluctuation of the environmental parameters, and such fluctuation is often useful for escaping local optimum. However, the vibrations can be removed using a filter before rendering if not suitable. Moreover, the amplitude can be varied during photo arrangement to obtain a better performance.

A performance evaluation confirms that the proposed algorithm provides about 60 FPS to show 700 photos with 512x320 pixels on a Windows 7 64-bit PC with Intel Core i7 CPU (3.20 GHZ), 12.0 GB memory, NVIDIA GeForce GTX 285. The system performance drops quickly when the number of photos exceeds 700. However, this is a reasonable number for photo browsing because of: (1) screen resolution (all photos are displayed on a screen and should be large enough to view individually); and (2) common number of photos belonging to an event. However, this number can be increased with different implementation (parallel run, multi-core CPU usage, etc.).

5 User Evaluation of D-FLIP

Photo arrangements with dynamic motions, shown in previous section, are expected to be effective in many situations such as viewing many photos at once, surveying a set of photos from different layouts, and finding pictures from the dynamic motions. Moreover, the smooth visual effect caused by interactions will keep up users' motivation to actively see and interact with photos. Our study investigates this further and explores whether it helps users perform browsing/sharing tasks easier.

We compared D-FLIP to the Windows Explorer (Explorer), a default Windows program. This is because although existing software (such as PhotoMesa, PhotoFinder, etc.) incorporate some of D-FLIP's animated properties, none of them supports animation of all photos in the collection. Instead, Explorer (of Windows 7) was chosen to be the most suitable candidate because: (1) it can be easily customized to have a separate area to store selected photos but still show the non-selected ones to support the story; (2) it includes many features of other photo browsing software such as: view all photos of the collection, sort or group photos (i.e. by name, tags, rating, etc.); and (3) it is a Windows built in software which is familiar and easy to use for users.

5.1 Task Design

The experiment is a modified version of [22] where a user shares a story with a friend by showing her photo collections. Each experiment session required two participants:

one narrator and one listener who sat beside each other in front of a 27-inch monitor (2560x1440 resolutions) displaying the narrator's personal photos. We measured task engagement from both narrator and listener to investigate the effect of the interactive application on the person who actually interacted with the program (the narrator) and on the person who only observed the interaction and listened to the story (the listener). Participants wore Emotiv EEG headset during the study.

There was a storyboard in D-FLIP to support the storytelling mode. The area had ten equal boxes which held ten selected photos (Fig. 6, left). In case of Explorer, we used two Explorer windows with one window above another. The bottom window was used to store 200 photos and the top window was used to store the selected 10 photos (Fig. 6, middle). In Windows 7, filenames, all bars and additional panes (e.g. Library, Details, and Navigation) were hidden to make Explorer comparable to D-FLIP in term of visualization and functions. Features of D-FLIP in this experiment included dynamic arrangement when overlaying a pointer, attraction for photos with similar colors. These features were chosen as they are comparable with Explorer.

5.2 Method

14 participants (9 males) between the ages of 19 and 32 volunteered for the study and were arranged into 7 pairs. The narrators brought to the experiment their 400 photos divided into 2 sets of 200 each. Photos were resized to 640x480, rotated if necessary by the experimenter before beginning the study.

Fig. 6. Storytelling mode with D-FLIP (left) and Windows Explorer (middle); An example of a narrator's task engagement changes in time in one session (right)

First, the narrator had adequate time to practice with D-FLIP and Explorer using a sample photoset (Fig. 6, left & middle). After this both users wore the Emotiv headsets. Each experiment session had two blocks each with either D-FLIP or Explorer. In each block, the narrator prepared a story in 5 minutes by selecting 10 photos from her first photoset. Then she told the story to the listener within 5 minutes. Both participants had a 2-minute break before the same procedure was repeated with the next program. NASA-TLX forms were given to both participants after each block.

In the storytelling step, we only recorded EEG signals from the listener. This was because the narrator needed to speak freely with facial, hand, and body movements which would contaminate the EEG signals. Narrators were instructed to use the software as normally as possible. This includes free using any desired features (including sorting, grouping, etc.). They were also encouraged to use the non-selected photos in the collection to enrich their story. In addition, all users were asked to sit comfortably.

5.3 Data Acquisition and Analysis

We used an Emotiv EPOC for measuring EEG signals. Its signals quality have been validated by previous studies (e.g. detecting and classifying Event Related Potentials

[16], evaluating visualization effectiveness [23], exploring the nature of decision making [4], and designing a BCI controlled video game [24]). It is portable, easy to setup, and most users can wear it comfortably for at least an hour [25].

We adapted the "sliding window technique" which is commonly used to process EEG signals (e.g. [26]). The recorded EEG signals (output at 128Hz) were segmented into 4s epochs with 2s overlaps between them. To remove artifacts, we adapted a decontamination process [20] with eye blink events detected by Emotiv SDK. Following this procedure, we removed totally 7.83% of collected signals in which 5.13% due to excessive signals and 2.7% due to high changes in combined power spectral.

Engagement index was calculated as in Pope et al. [21]. β, α, and θ, are the combined power in the ranges of 13-22Hz, 8-12Hz and 5-7Hz frequency bands: *Engagement Index* $= \beta / (\alpha + \theta)$. We used F3, F4, FC5, FC6, P7, P8, O1, and O2 channels as they are the surrounding channels of Cz, Pz, P3, and P4 which were used by Pope et al. [21]. Engagement indexes were calculated in 4s window for each task & user separately and then normalized for the narrator or listener.

6 Results and Discussion

Table 1 summarizes the results, which are the averaged task engagement during the session, for two types of tasks (Preparation/ Telling), two types of programs (Explorer/ D-FLIP), and with two participant types (Narrator/ Listener). Fig. 6 (right) shows a time series sample of engagement during the preparation task of one narrator. Fig. 7 (left) shows the task engagement values of narrators in each experiment session between two programs. T-test shows that there is a significant difference between those two ($t = 2.95$, $p < 0.05$). However, we found no significant differences between two types of program in listeners ($p > 0.5$). The details are shown in Fig. 7, middle (Story Preparation) and Fig. 7, right (Storytelling). This may due to the individual differences where the listeners did not directly interact with the programs hence the measured task engagement depended more on the storytelling skills of the narrators.

Table 1. Averaged task engagement for narrators and listeners

Tasks	**Story Preparation**		**Story Telling**	
Programs	Explorer	D-FLIP	Explorer	D-FLIP
Narrator	0.603	0.554		
Listener	0.480	0.436	0.512	0.440

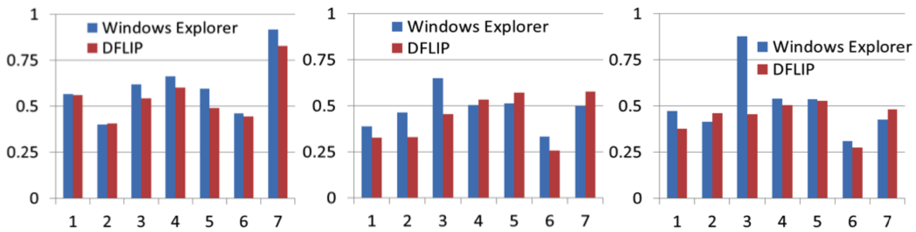


Fig. 7. Task Engagements of narrators in Story Preparation task (left); and listeners in the Story Preparation task (middle) and Story Telling task (right).

For the NASA-TLX, only mental demand (for narrator) presented with a significant difference ($p < 0.05$) between two tasks: Explorer (mean: 9.14) and D-FLIP (mean: 5.00). For the listener, there was no significant difference ($p > 0.05$) for any of the NASA-TLX questions. These results showed that the narrators had higher engagement when using the Explorer compared to D-FLIP. This implies that narrators need to put in more effort with Explorer than D-FLIP to complete the same task in the same amount of time. Consequently, besides the benefit of having more interest, D-FLIP makes the task easier to perform. The results of NASA-TLX (for both participant types) are consistent with the task engagement results.

Users had adequate practice time with D-FLIP. However, the familiarity with Explorer may result in a more skilful performance. Hence, the actual difference between measured task engagements of two programs might be larger if a less famous program (e.g. PhotoMesa) was used instead of Explorer. Our results show that even if there are effects of familiarity, D-FLIP still requires less task engagement and mental demands.

7 Discussion

An interesting finding from our user evaluation is that D-FLIP requires less mental effort and task engagement compared to Explorer in narrators but not listeners. A probable cause is that only the narrators physically actually interact with the programs. Hence, with less task engagement or workload in performing the task, interacting with D-FLIP is easier and more pleasurable.

Several factors of D-FLIP may contribute to this result viz.: high visibility with variety of layout, dynamic and smooth motions of photos, gathering photos based on similar attributes. Consequently, users can focus on interactions with other users and the story contents with less effort in performing the task. Additionally, D-FLIP keeps motivation and interest in users due to the dynamic, flexible, and interactive motions of photos produced by the proposed principles, thus making D-FLIP a particularly powerful system for visualizing large collections of images.

Our next step is to improve the algorithm categorizing the parameters and optimizing it for different types of interactive applications. Besides existing content types (i.e. text, broadcasts, movies, etc.), our algorithm can adapt to work with new and emerging content types such as dreams visualization [27] where values of parameters from fMRI patterns for an image. This can also help to build an ecosystem of photos which includes various promising features such as pigeonholing, printing, automatic acquisition of meta-data, evolving into photo sharing sites, and coordinating with social network services. We can also explore other interaction devices such as multi-touch digital tables, voice recognition, brain-machine interface, and other sensing devices to further enhance the fluidity of interaction in different application contexts.

Although the focus of this paper is not on a novel evaluation methodology, we believe that our way of measuring task engagement using EEG offers greater insights into workings of an application. Our measured task engagement is consistent with the NASA-TLX results; providing a source of external validity to our measurement mechanism. It can also be improved to capture emotions (e.g. relaxation, meditation) and other users' inner states (e.g. error awareness).

8 Conclusion

We presented D-FLIP a program which displays a set of photos flexibly, dynamically, and interactively. The underlying algorithm adjusts the displaying principles adaptively when users interact with the program. Our performance evaluation shows it can handle smoothly at least 700 photos. Our user evaluation shows that D-FLIP requires less task engagement and mental effort from users allowing them to enjoy the content rather than manage it.

References

1. Kirk, D., Sellen, A., Rother, C., Wood, K.: Understanding photowork. In: CHI (2006)
2. Bederson, B.B.: PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps. In: UIST (2001)
3. Dachselt, R., Frisch, M., Weiland, M.: FacetZoom: a continuous multi-scale widget for navigating hierarchical metadata. In: CHI (2008)
4. Graham, A., Garcia-Molina, H., Paepcke, A., Winograd, T.: Time as essence for photo browsing through personal digital libraries. In: JCDL (2002)
5. Ahern, S., Naaman, M., Nair, R., Yang, J.H.-I.: World explorer: visualizing aggregate data from unstructured text in geo-referenced collections. In: JCDL (2007)
6. Gomi, A., Itoh, T.: A personal photograph browser for life log analysis based on location, time, and person. In: SAC (2011)
7. Brivio, P., Tarini, M., Cignoni, P.: Browsing Large Image Datasets through Voronoi Diagrams. IEEE Trans. Vis. Comput. Graph. 16(6) (2010)
8. Itoh, T., Yamaguchi, Y., Ikehata, Y., Kajinaga, Y.: Hierarchical data visualization using a fast rectangle-packing algorithm. IEEE TVCG 10(3) (2004)
9. Gomi, A., Miyazaki, R., Itoh, T., Jia, L.: CAT: A Hierarchical Image Browser Using a Rectangle Packing Technique. In: Book CAT: A Hierarchical Image Browser Using a Rectangle Packing Technique (2008)
10. Rother, C., Bordeaux, L., Hamadi, Y., Blake, A.: AutoCollage. SIGGRAPH (2006)
11. Wang, J., Quan, L., Sun, J., Tang, X., Shum, H.-Y.: Picture Collage. In: CVPR (2006)
12. Rother, C., Kumar, S., Kolmogorov, V., Blake, A.: Digital Tapestry. In: CVPR (2005)
13. Kim, S., Yoon, M., Whang, S.M., Tversky, B., Morrison, J.B.: The effect of animation on comprehension and interest. JCAL 23(3) (2007)
14. Rieber, L.P.: Animation, incidental learning, and continuing motivation. Journal of Educational Psychology 83(3) (1991)
15. Kintsch, W.: Learning from text, levels of comprehension, or: Why anyone would read a story anyway. Poetics 9 (1980)
16. Vi, C., Subramanian, S.: Detecting error-related negativity for interaction design. In: CHI (2012)
17. Hagemann, D., Hewig, J., Walter, C., Schankin, A., Danner, D., Naumann, E.: Positive evidence for Eysenck's arousal hypothesis: A combined EEG and MRI study with multiple measurement occasions. J. of Personality and Individual Differences (2009)
18. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. ACEM 78(5) (2007)

19. Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., Grier, R., Warm, J.S.: 'Fundamental Dimensions of Subjective State in Performance Settings: Task Engagement, Distress, and Worry'. Emotion 2(4) (2002)
20. Berka, C., Levendowski, D.J., Olmstead, R.E., Popovic, M.V., Cvetinovic, M., Petrovic, M.M., Davis, G., Lumicao, M.N., Westbrook, P.: Real-time Analysis of EEG Indices of Alertness, Cognition, and Memory with a Wireless EEG Headset. IJHCI 17(2) (2004)
21. Pope, A.T., Bogart, E.H., Bartolome, D.S.: Biocybernetic system evaluates indices of operator engagement in automated task. Biological Psychology 40 (1995)
22. Rodden, K., Wood, K.R.: How do people manage their digital photographs? In: CHI 2003 (2003)
23. Anderson, E.W., Potter, K.C., Matzen, L.E., Shepherd, J.F., Preston, G.A., Silva, C.T.: A user study of visualization effectiveness using EEG and cognitive load. In: EuroVis (2011)
24. Ames, M.G., Manguy, L.: PhotoArcs: ludic tools for sharing photographs. In: MM (2006)
25. Ekandem, J.I., Davis, T.A., Alvarez, I., James, M.T., Gilbert, J.E.: Evaluating the ergonomics of BCI devices for research and experimentation. Ergonomics 55(5) (2012)
26. Lee, J.C., Tan, D.S.: Using a low-cost electroencephalograph for task classification in HCI research. In: UIST (2006)
27. Horikawa, T., Tamaki, M., Miyawaki, Y., Kamitani, Y.: Neural Decoding of Visual Imagery During Sleep. Science (2013)