Flutter: Directed Random Browsing of Photo Collections with a Tangible Interface

John Williamson
Dept. Computing Science
University of Glasgow
Glasgow, Scotland
jhw@dcs.gla.ac.uk

Lorna M Brown
Microsoft Research
Cambridge, UK
lornab@microsoft.com

ABSTRACT

Large collections of photographs are commonplace, and many interfaces for viewing, sorting and organizing them have been proposed. This work describes the design and implementation of a "living photo frame" – designed not to navigate or browse collections but to create an enjoyable activity from a collection of images. Tangible interactions with a tablet-style PC are used to bind the user closely to the system. Every interaction is logged and used to gradually evolve the structure of photo collections.

Categories and Subject Descriptors

H.5.2 [**Information Systems: User Interfaces**]: Auditory (non-speech) feedback; H.5.2 [**Information Systems: User Interfaces**]: Interaction Styles
; H.5 [**Information Systems**]: Miscellaneous

General Terms

Human Factors

Keywords

photo, browsing, Monte Carlo, tangible, inertial

1. INTRODUCTION

It is now common for people to have extremely large collections of personal digital photographs; the profusion of mobile devices with imaging capabilities means that large segments of the population take photographs on a regular basis. The technologies facilitating capture of images have outpaced the design of technologies for interacting with the resulting collections. Traditional methods originally intended for chemical photographs – such as albums, photo frames, or simple printed images – are still widely used. Although these are well-proven and effective, they do not always support the ways in which a user may wish to engage with a collection of images. Many advanced browsing and searching interfaces have been created; however these are often concerned with efficient achievement of well-specified goals – not something commonly occurring when simply browsing through an album for fun. This paper describes a system designed to support casual interaction with photograph collections, employing a simple tangible interface combined with an online recommendation system. Rather than facilitating directed searching tasks, Flutter is designed for contemplative, background interactions; simply sitting and enjoying a collection of images. The system is intended to delight and surprise the user, taking as its inspiration the fun of sifting though a pile of photographs lying on a table, but extending the interaction to scale it to the magnitude of contemporary digital photo archives.

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, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
DIS 2008 Cape Town, South Africa
Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...$5.00.

2. DESIGN AIMS

The original design goal was to create an interactive system that could take a large set of disorganized digital photographs – typical of the collections many people have – and somehow engage the user in interacting with them. The target group for the design was home users, who wish to interact with their own collections of personal photographs, potentially with multiple users sharing a common collection of images (such as in the case of shared family photographs). One of the key goals is that the user should be tightly bound into the interaction, which should be simple and immediate. The responsiveness and simplicity of the interface are vital in making use enjoyable. Making the system a source of pleasure is emphasized over the enabling of specific actions. As a consequence, the focus of this design was on using basic tangible interactions, without layers of hidden complexity, rather than conventional graphical interfaces.

The initial investigations focused on determining what kinds of photographs people have; what tasks they perform with them (and what tasks they would like to perform but cannot); and what structure collections have. These subjects have been widely studied in the literature ; in particular [6] examines the use of photographs in the home, based upon a comprehensive study, while [10] presents a detailed study of the use of personal digital photographs. Both noted the general lack of organization of digital photographs, and the use of very simple exploration techniques. Complex searching activities were not found to be of particular benefit to users dealing with their personal archives. Rodden et. al. in [17] also examined digital photograph activities, observing a distinct lack of annotation activity and the utility of temporal structuring in exploration of photo archives.

2.1 Structures and Orderings

From a technical perspective, the structure of photographic collections – the metadata which relates them to each other – is of great importance. If the system is to engage the user at anything other than the most trivial level, the presentation of images must, in some way, be sympathetic with the way in which people perceive

the relationships between photographs. Without this, the best that could be hoped for is interaction which appears completely random to the user.

Unfortunately, many of the attributes of a photograph which are most important to people – that is those attributes by which a typical user might describe a class of images – are unavailable to photographic systems. Laborious manual annotation can introduce some of this structure, but only insofar as annotators have sufficient insight, and motivation, to describe images in terms of their future possible use. Many directed queries upon photographic collections involve relationships between people, especially the presence of others in the photographs (or indeed outside of photographs – an image from so-and-so's home, for example). Such information cannot practically be extracted automatically (although some systems, such as [11] and [18] attempt to use facial matching to identify the presence of individuals). Other queries involve high-level "events": holidays, celebrations, and gatherings. Some of this can be extracted from the available temporal structure of photographs; images can at least be clustered into temporally dense blocks which are likely to relate to specific events. However, the identity of such events can generally only be identified by the photographer or other participants.

2.1.1 *The Richness of Temporal Orderings*

One piece of metadata which is almost always available is the date and time, as most digital cameras and camera phones timestamp each photo (both on the file itself, and embedded within the EXIF data), either on capture or on upload. Although timestamps are the most humble of photographic metadata, their apparent simplicity conceals a richness of structure. Photographs are generally clustered in time – it is highly unusual for someone to take photographs at a constant rate. The clusters that result often delineate meaningful events; a dense set of images might been seen during a celebration, for example. Longer term "events" such as holidays or trips consist of multiple dense clusters which are unusually close together. Drucker *et al* [5], Girgensohn *et al* [7] and Platt *et al* [15] discuss temporal clustering algorithms for organizing photograph collections.

By considering not just the clustering of timestamps along a standard timeline, but also the aggregations which form when timestamps are organized modulo some period, even more structure can be ascertained. A yearly "projection" of photographs is likely to show clusterings around those times when holidays are most common, such as summer or Christmas. A weekly projection distinguishes weekday events from weekend events. A daily projection separates evening events from daytime ones. Temporal information relates photographs in a collection to *each other*, and thus is ideally suited to exploration tasks, where the interrelationship of elements is required.

2.1.2 *Content-based Attributes*

Attributes can also be extracted automatically from images; see for example [12] or [4]. However, such attributes rarely correspond closely to the structure humans would assign to the images; few people will remember that two photographs have similar colour casts or similar textures. Most content-based attributes are based on colour information, either aggregated (e.g. as histograms) or as direct transformations of image pixels. While there is some evidence arrangement according to these attributes can be beneficial ([16]), and some gross classes can be automatically distinguished (portraits versus landscapes, indoor versus outdoor or the presence or absence of humans) they offer little in comparison to other available metadata.

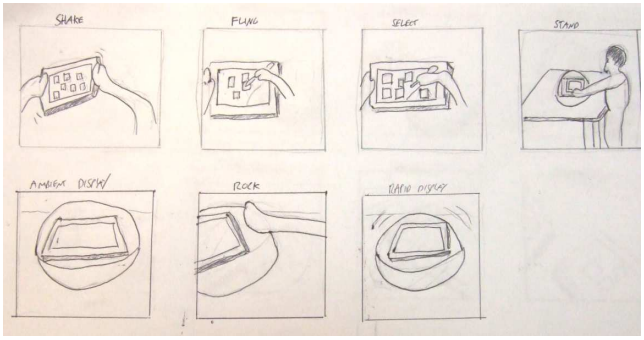

Figure 1: One of the original storyboards, showing the initial "rocking-bowl" interaction.

2.1.3 *Tagging and Annotation*

Data can be manually added to images, by attaching tags or more detailed notes. Tagging is popular, as websites such as Flickr demonstrate. However, tagging is time-consuming and tedious, and few users spend the time to tag their complete collections [10]. The consistency of tags is important for their later use for recall, but it is difficult to be consistent in the application of tags when many thousands of images are involved. Tags are better suited to specific searching tasks than exploration activities, as they define relationships between photographs and external objects, rather than relationships between photographs themselves.

2.1.4 *Geotagging and other sensed data*

Many cameras are now found as part of phones, or can interface to other sensing technologies (e.g. via Bluetooth). This opens the possibility of increasing the metadata recorded when a photograph is taken. Of particular interest is geotagging, where the location of the photograph, obtained from GPS, cell data (when the camera is embedded in a mobile phone), or other sources, is attached to the image. If heading can be obtained from an electronic compass, this can also be logged, providing a record of the way the camera faced along with its location.

Carper et. al. [14], Gurrin et. al [9] and Naaman et. al. [13] discuss the impact of geotagging. Because location of photographs is an attribute which people can directly relate to (and one in which photographs are often described in terms of), such data is of particular value; Bentley and Metcalf [3] discuss some of the ways in which people relate to georeferenced photographs. However, it is still largely in the experimental stages, and few commercially available devices support geotagging out of the box.

2.2 Scenario Development

The results from the literature prompted the exploration of a number of scenarios for pleasurable interactions with photo collections. These scenarios were storyboarded (see Figure 1 for an example) and then analysed at a small design workshop.

It was decided to investigate tangible interactions using some type of frame-like device; the constraints on form factor were largely derived from the technical and physical limitations of the implementation hardware. The device must to be large enough to display pictures clearly, while being small enough that it can be picked up, shaken and manipulated by the user. Given these constraints, physical dimensions around that of a traditional photo frame seems appropriate. The familiarity of digital photo frame-style devices to many people was also considered an advantage.

A further outcome of the design workshop was the decision to

focus on contemplative, pseudo-random scenarios of use, where users engage in semi-passive interaction with the collection. It is clear, from the review of the literature, that the lack of metadata attached to images – and the irrelevance of much of it to users – can pose challenges in presenting interesting images to users. This suggested that structure might better be seeded with captured metadata (e.g. timestamps) and then refined through automatic observation of user interaction with the device. By seeding the structure based on user interactions, the system evolves with use, becoming a "living frame".

3. DESIGN OF FLUTTER

These initial decisions led to the development of a number of concepts which underlie the design. These are outlined below.

3.1 From Collections to Activities

The fundamental purpose of the design is to produce an artifact which takes a collection of images, with some structural information, and transforms it into a pleasurable activity. The object of the system is to maximize a user's interest in the media which are displayed; each image should seek to provoke a response. Additional structure accretes as interactions extract evidence from users about their interests in images.

3.2 Supporting Ignorance

An individual picking up an interactive photo display often does not have a clear idea of what images he or she wishes to see. This partially explains why many sophisticated and powerful organization and query interfaces are not widely adopted. Few users know what they want to see before they begin; fewer still are able to distill those intents into meaningful queries across the attributes of images which the system observes. Photo journalists, archivists or other workers with very specific and well-defined needs may benefit from such interactions. This use case, however, is exceedingly rare among home users exploring personal photograph collections. Indeed, in the design considered here, it is important that users be willing to relinquish direct control over what they see, in return for the enjoyment of surprising interactions.

3.2.1 Iterative Refinement of Belief

Although users may not have a definite idea of what images they would be interested in seeing, or are unable to communicate their preferences given the available metadata attributes, they may instead be able to iteratively refine selections to find images of interest. The presentation of a sample from a large set of images can stimulate memories; users can then follow paths through photo space by indicating that they would like to see more images "similar" to one or more of those displayed. Using rich similarity metrics is essential in obtaining effective navigation by this means.

This style of interaction has much in common with Bates' [2] "berry picking" model of information retrieval. In this model, users wander through an information space, finding results and modifying their queries as they go. The final goal of the user adapts as they bounce through the results from each previous query.

3.3 The Gradual Etching of Structure

At the outset of an interaction with a collection of photos, the only information the system can use to select images is derived from the metadata attached to (or extracted from) those photographs, along with some pre-defined function which determines how those metadata elements should be combined to best select images. This information, while useful, is unfortunately completely impersonal. One of the aims of the design is to transform the collection of photographs into a personalised, meaningful artefact; the interactors must be able to stamp their identities onto the archive. Rather than having a tedious tagging or album formation process as photographs are added (a process that is considered an unpleasant chore [6]), Flutter attempts to extract information from the patterns of use of users, and from this refine the functions which drive the selection of images. Learning of user interest in regions of photographs has previously been explored in [20], where browsing activity was used to automatically determine which areas of an image are of most relevance.

3.4 Interaction to Ambient Display

Two common styles of interaction with existing physical photo collections are permanent display (in the form of framed images) and private viewing or contemplation (for example, flicking through an album). The design is intended to support both of these styles, but also to allow information to flow between them, so that the nature of the display in its background, ambient mode is affected by the interactions of users in the active exploration mode. The selection of images shown in ambient mode is guided by the interest expressed in them by previous users.

4. SCENARIOS OF USE

The device is placed in a household in a public area, such as a lounge, where a number of people regularly interact with it.

A user picks up the device, removing it from the stand. He gives it a gentle shake, and a couple of pictures fall onto the surface. Nothing of interest appears, so he shakes again, until he notices a long-forgotten holiday image. He zooms it up to fill the screen, and then pokes at it to reveal other photographs taken at a similar time. Seeing several he would particularly like to display, he pushes them to one side of the screen where they fall into a "storage bin". He then replaces the device in the stand and leaves it be. The images he placed in the storage bin are mixed into the gradually fading ambient display. The system remembers that the first few images were ignored, and that special emphasis was placed on the images which were put on display; it then reweights the image scores (6.1) so that future image selection can incorporate this information.

A second user – bored and looking for some activity to occupy her, picks up the device and sits down with it. She shakes the device and observes the photographs fluttering down. Those of no interest she throws to the corners of the screen, while arranging the remaining ones around the centre. The system selects photographs based on how recently they have been viewed, in combination with how much interest previous users have given the images. As a consequence, many images which she has not seen for some time appear, along with those that other members of the household have found particularly interesting.

A third user is interacting with the device and finds two photos of her friend. As they were taken on different occasions, they are not grouped together. She encircles the two images, thus indicating to the system that these images are related. Flutter remembers this and will now group these two photos as being related in future interactions.

5. DIRECTED RANDOM BROWSING

The fundamental basis of the design is that photographs displayed should be of interest to the individual who is exploring the collection. "Interestingness" is clearly not an accessible variable. The property arises as a joint function of the interactor's mind and the image itself. The state of a user's mind is of course hidden from the display, and only indirect evidence can be accumulated to guide

selection towards those which will be most effective.

The state of even a single user's mind varies rapidly over time and will be strongly influenced by the previous actions of a display. The display process is not stationary; dynamic effects are critical components. Continuously assessing the likely interest of each individual image is hopeless. Instead, approximations can be chosen which employ carefully balanced randomness to stimulate the viewer. Furthermore, the response of user can be continuously fed back into the selection mechanism, so that indications of salience or non-salience can be observed and used to shape the future organization of collections.

5.1 Balancing Entropy and Filtering

It is arguable that the most interesting arrangements lie between the extremes of utter disorder (which without any discernible meaning is seen as a uniform sandblast of noise) and perfectly rigid ordering, which is so predictable as to hold little interest once the underlying pattern is inferred. The design here follows such a philosophy, drawing images in a pseudo-random order, but directing the sampling towards those which are estimated to be of greatest interest. The entropy of the sampling injects surprise; the bias of the selection injects structure. A balance must be struck between these competing forces to effect the most compelling displays. The "salience functions" described later formulate this balance as a term in a simple equation defining the sampling over images.

5.2 Persistence of Interaction Data

The structure of the system gradually increases as users interact with it. This structure is determined by the history of interactions with images. Each time an image is displayed, zoomed, moved, grouped or destroyed the event is stored in a database. When image selection is being computed, the log for each image is analysed, and the salience metric uses the recorded data to calculate the overall score for that image.

6. THE DESIGN OF SALIENCE METRICS

The "flow" of photographs is determined by the salience metric which scores each image; good design of this metric is critical in making the display captivating. The function of the metric is to combine all metadata, including data recorded at the moment of capture, and evidence subsequently acquired from interactions, into a single number which represents how interesting the photograph is likely to be.

6.1 The Aim of a Salience Metric

Each photo that is displayed should, in the ideal case, be the one which is most interesting to the user at that particular point. The salience metric quantifies an estimate of the "interestingness" of an image. These quantifications must by necessity be relatively crude. However, they seek only to introduce sufficient regularity into the organization of images that the user's interest in the display is maintained.

Quite a number of factors can provide useful proxies for the potential salience of a photograph. These can either be prescribed by a designer (as in the present implementation), or users could tweak the weightings to correspond to their current mood and desires. Ascertaining these factors is difficult; however some potential basic factors are:

- **Least Recently Viewed** Score highly those images which have not been seen for a long time. This will tend to bring to the fore images which may have been forgotten.
- **Most Frequently Interacted With** Score highly those images which provoke a sensed response from a user; for example those that have most often been examined in detail.
- **Viewed by Most Distinct Individuals** In scenarios where multiple parties may interact with the display, and where those parties may be uniquely identified, images may be scored by how popular they are among the user group. This might, for example, be a worthwhile metric when the display is used in a family environment. A simple variation empahsises images which are most viewed by individuals other than the current viewer.
- **Most Distant from Images Recently Displayed** The image similarity metrics described in Section 2.1 can be used to score images such that photographs which are most dissimilar to those currently or recently shown are most likely to be shown now. The inverse of this (those most similar to those shown) might also be used, but in the displays described in this paper, there are specific mechanisms for exploring clusters of related photographs (see Section 8.4).

7. DESIGN SUMMARY

From these decisions a series of working initial prototypes were created, and the design of the interface was refined iteratively until the design was finalised. Each of these prototypes was informally evaluated with a number of people, primarily to determine whether the interface evoked a positive emotional response, as well as to eliminate any serious usability flaws. The final feature set, interaction methods and feedback design of the ultimate prototype are described in the following sections. Images of the final implementation are shown in Figure 2.

7.1 Features

The implemented system functions in two distinct modes; interactive mode and ambient mode. The mode changes when the device enters or is removed from its stand. In interactive mode, the display appears as a surface upon which photos can be dropped onto and then arranged. The following features are available:

- Photographs can be introduced, according to the random sampling scheme.
- Photographs can arranged, and zoomed for closer examination.
- Photos can be manually grouped, introducing new relationships between images.
- Once introduced, images gradually fade away and then disappear, unless they are in some way interacted with.
- Images can be stored in a "storage bin" so that they persistent between sessions, and so that they appear in the ambient mode slide show; conversely they can be placed in a "kill area" so that they very rapidly decay to nothing.
- Photographs related to a currently visible image can be explored, by "shaking out" related images.

In ambient mode, the display works like a traditional digital photo frame. Images are randomly selected from the set of images lying in the storage bin. They are displayed for a few minutes, then the display crossfades to a new image. If the device is moved while in ambient mode, the image transition occurs immediately.

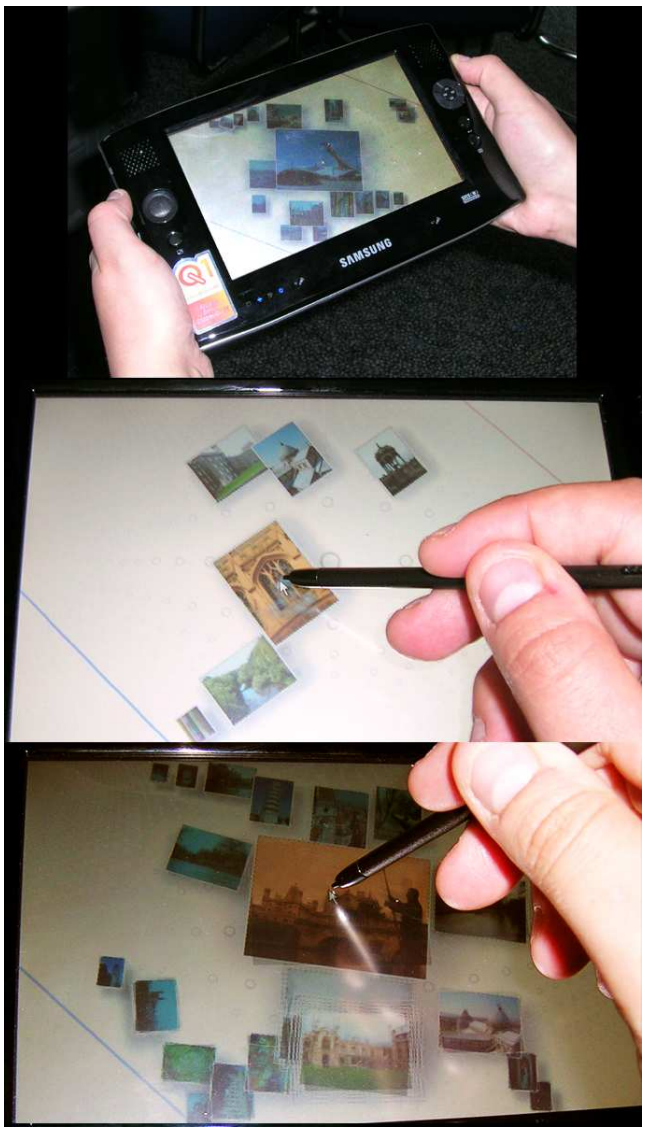

Figure 2: The final implemented system, running on a Samsung Q1 UMPC. The visual design is bare, with only a subtle background to indicate the distortion of space. The photographs are framed and shadowed to improve separation between them. The regions at the top right and bottom left corner, are the "storage bin" and "quick kill" area respectively.

8. INTERACTIONS

The interaction with the display is intended to be rich, simple and grounded in physical metaphor. By using and extending simple physical interactions, clean and intuitive controls can be introduced. The display is designed like a table upon which images can flutter down and then be quickly arranged, examined and grouped if so desired.

There are two primary forms of input used in the system; inertial sensing, which senses shaking for the basic action of introducing new images; and touch-screen interactions, which are used to closely examine and loosely organize images once they have been introduced.

8.1 Inertial Interactions

Shaking the whole device back and forth introduces new images from the collection. Each image flicks out and drops down onto the virtual surface of the display, landing in the centre. Continued shaking creates a stream of images dropping down onto the surface. The shaking is detected using the accelerometers of the SHAKE device (see Section 11). Only the z-axis is used; this is the axis which is normal to the plane of the display. The signals are high-pass filtered, and then drive a simulation of a spring-mass-damper system. When the mass at the end of the spring crosses a preset threshold, an introduction event is generated, and a new photo floats down.

All of the images (excepting those placed in the storage bin) can be removed from the display by inverting the entire device, as if dropping the images off from the display. After a short time in this orientation (to avoid accidental removals), the images are removed. Since it is highly unlikely that the device will be placed and held in this orientation, as the screen will be entirely obscured, this motion is generally robust to accidental activation.

The accelerometers attached to the display, as well as providing control inputs, also measure the orientation of the device with respect to gravity. This is used to rotate photographs so that they always remain in the upright orientation as the device is rotated. This enables easy sharing of the display between co-located users, as when the device is tilted towards another person the photos tilt with it, positioning them correctly for the person viewing them.

8.2 Stylus Interactions

The stylus interactions (see Figure 3) are relatively simple. Images can be tapped to bring them to the centre and zoom them; they can be dragged around to place in different areas of the screen; they can be dragged into the storage bin placed at one corner where they will remain persistently, or to the kill area where they rapidly fade away; and the user can draw outlines around a collection of images to join them together into a group. Photos in the storage bin or kill area can be removed simply by dragging them back out. Because these regions are placed in the corners, where the visual distortion (9.2.1) is high, there is a great deal of room for photographs in these areas. Groups of images, once created, are linked to each other by stiff spring connections so that the photographs move as a group. The group information is used in the calculation of tangible photo clusters (Section 8.4).

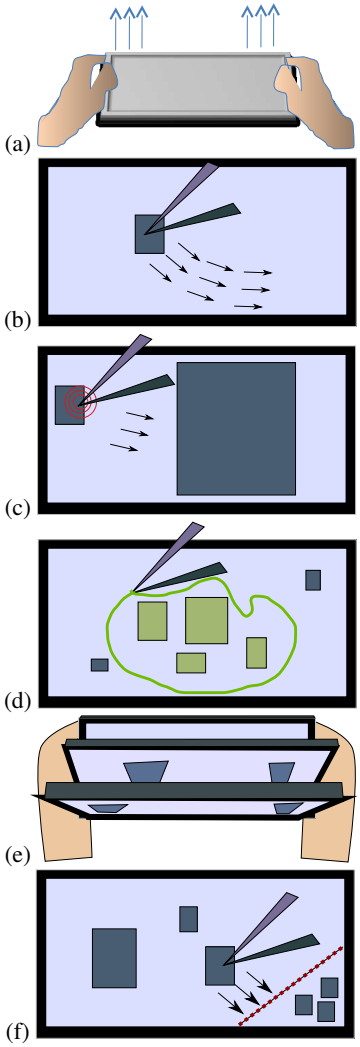

Figure 3: Interactions. (a) Shaking the device to introduce new photographs. (b) Dragging photographs around to rearrange (c) Tapping a photograph to zoom it. (d) Dragging around a number of photographs to group them. (e) Flipping the device to dump clear the surface. (f) Placing images in the storage bin, to be preserved for future sessions, or placing them in the quick kill area to eliminate them.

8.3 Ambient Interactions

When the device is in ambient display mode (resting in its stand), the display is completely occupied by a single image. After a set time, this fades to another image, selected from the set of images lying in the storage bin. Gently prodding the device will prompt it to move on to the next image. Since the stand on which the display rests can rock (see Section 11.2), pulling one side down will result in a rocking motion which will cause the display to run through a few images before stabilizing.

8.4 Tangible Photo Clusters

The system supports interactive "drill-down", as discussed in Section 3.2.1. This is implemented as a "tangible cluster" metaphor, where related photographs form a cluster around an original image. These can be "shaken out" by stimulating the original image. This process involves realistic feedback from a physical model, which rapidly communicates the quantity of photographs in a cluster.

The distance between each pair of photographs given some metric is computed; in the implemented system the time-difference of the images is used along with the grouping information added by users. Each photograph is then assigned a neighbourhood of images for which this metric is smaller than some threshold. When only the timestamp is used, for example, this results in photographs having a local neighbourhood set of images taken within a certain period of time. This is intended to capture clustering of photographs around important events. When further data has been gathered through user interactions with the system (e.g. by users grouping activity), additional relationships will be present, e.g. the photos may feature the same person, or be taken in the same location.

Users can explore the cluster of images associated with a particular photograph by dragging it with the stylus and then performing a gentle shaking motion with the stylus. The images within the cluster are associated with individual simulated masses attached with springs to the original photograph object. When excited by the dragging motion of the original photograph (Figure 4), the masses bounce around, striking a virtual container around the original image. These impacts produce audio and vibrotactile feedback, like balls rattling in a box, in a similar manner as the Shoogle system [19]. The kinetic energy of these impacts is accumulated, and the result of this accumulation is linked to the gain of the feedback, so that the intensity increases as the user continues to excite the image.

Once this integrated value exceeds a threshold, after several seconds of excitation, the impacts induce the introduction of the image associated with the impacting mass as it strikes the boundary. This causes the related photographs to begin to "fall out" of the original image as it is shaken. If stimulation is maintained, every related image will eventually be shaken out.

9. FEEDBACK DESIGN

The feedback from the system strongly affects the quality of the interaction. In the design discussed here, the responsiveness and liveliness of the interaction are critical; the timing and dynamic response of the system need to be tuned so as to be as immediately engaging as possible. Utilising the non-visual modalities is important in this design, both because the visual focus should be on the photographs and not the interaction, and because high-quality, physically-motivated audio and haptic feedback greatly enhance the tangibility of the interface.

9.1 Visual Display

The display is designed to be as uncluttered as possible, with no GUI controls visible at any time. All interactions involve either di-

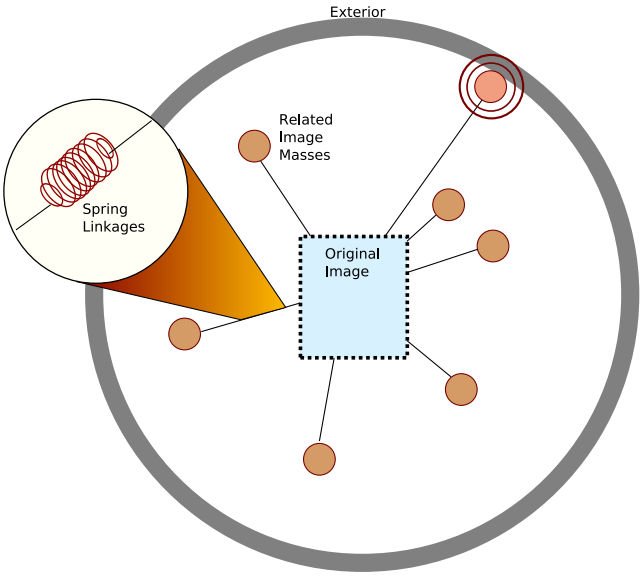

Figure 4: The tangible photo clusters. A physical simulation of masses attached to the original image by springs is used; when these masses strike the exterior boundary, feedback events are generated.

rect stylus manipulation of the images or whole-device movements. Photographs are framed and shadowed to make clear the ordering of images and avoid clashing when similar coloured photographs overlap. The visual appearance is intended to be as simple and clean as possible, without distracting elements, making the photographs the focus of attention. The layout of photographs – as a pile of slightly overlapping images – is informed by the studies of Grant et. al. [8].

After photographs are introduced, they begin to age, and gradually fade to black over a period of minutes. Once wholly faded, they are removed from the display entirely. If an image is interacted with in any way (tapping, dragging, etc.), its "age" is reset. The removal of images is necessary to avoid clutter as streams of images are introduced. The fading mechanism serves to inform the user of the impending removal of images.

Figure 2 shows the appearance of the final prototype.

9.2 Physical Modeling: Potential-based Models and Realistic Animation

All of the interactions with the system are model-based, with potential field models driving the motion of objects and the production of feedback in response to input movements, in a similar manner to BumpTop ([1]). This leads to smooth natural animation, and makes linking the audio and vibrotactile events to the visual interactions simple. Actions such as tapping to zoom create time-varying potential fields which drag the image to the centre of the screen and pull it "up" towards the camera. The result is a smooth, clean animation as the photograph rises up, pauses for a few seconds, and then falls back down. Dragging with the stylus creates a spring linkage between the stylus and the image the stylus went down on, allowing natural flinging motions (Figure 5). The automatic layout of images is also solved using a potential field model. Repulsive forces between images, whose strength varies as the area of overlap, arrange images so that the overlap function is minimized. The direction of the force is aligned so that photographs are always propagated outward, away from the centre of the screen.

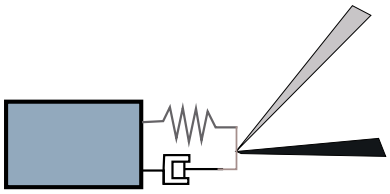

Figure 5: When dragging images, the motion of the photograph is driven by a spring-mass-damper system, with the other end of the spring linked to the current stylus point.

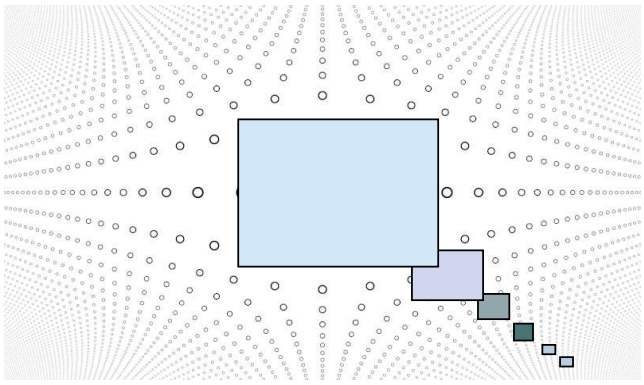

Figure 6: The distortion used in Flutter. The circles show the scale at each point on the screen. The scale of images rapidly reduces towards the corners, where it gradually levels out.

This causes them to naturally shrink out towards the corners, where most space is available.

Although not implemented in current versions, the physical modeling basis for the animation opens up the possibility of linking content to physical parameters; for example, making older feel images slightly "gritty" (by introducing random fluctuations into the friction function) or creating the impression of weight in images which are grouped with many others.

9.2.1 Space-extending Distortions

The display features a fisheye-like distortion, which scales images according to their distance from the centre. The geometry of the images is not distorted; the images are simply scaled according to the distortion function, evaluated at the centre of the image. The display is designed to resemble a pile of images, with a few large, high-resolution images lying at the very top, and many smaller thumbnail images towards the edges. Figure 6 illustrates this. This allows overviews of a large number of images, while still retaining detail on the few images at the top of the pile.

9.3 Audio and Vibrotactile Feedback

Many of the interactions for manipulating photographs (such as shaking or tapping on an image) temporarily obscure part of the display. Vibrotactile and audio feedback are provided along with visual responses, to mitigate some of the effects of the obstruction of view, and to enhance the tangible nature of the interaction. Audio and vibrotactile events are paired together and synchronized. The audio design is based around metaphorical auditory icons rather than abstract sounds. For example, the sound of cards being flicked on to a table is used for the introduction of new images; the sound of marbles rattling for the shaking out of related photographs from a cluster; and a gentle slurping sound is used when photos are

dropped off the display when it is inverted. The vibrotactile events follow roughly the amplitude contour of the sounds they are attached to, with short, weak impulses for the introduction of photographs, and a long gradually increasing vibration for the slurping of images being dumped off the device. These sounds are designed to be unobtrusive and to be natural consequences of the motions that invoke them.

10. IMPLEMENTATION DETAILS

10.1 Interestingness Sampling

The underlying algorithm is simple Monte Carlo sampling from a discrete set. At the introduction of each image, samples can be drawn from a *salience distribution* $P(i)$. This assigns a probability of selection to each image i. This function can be obtained by scoring each image according to a combination of metrics, and then normalizing the result such that $\sum_i P(i) = 1$. Ordering the images arbitrarily, and computing the cumulative distribution function $C(i) = \sum_{k=0}^{i} P(k)$, sampling can be performed by drawing a random number r uniformly on $[0, 1)$ and then iterating through each image until the first photograph for which $C(i) \geq r$ is satisfied.

10.2 Mixtures

The computation of the salience score for an image is derived from a combination of the individual factors described in Section 6.1. A simple linear weighting of the available scores, plus a bias term which defines the "background randomness" of the display suffices. The background term approximates the selection. Each image i then has a score $S(i) = \sum_{j=0}^{n} \alpha_j S_j(i) + \beta$, where the α_j's are the weighting for each individual metric $S_j(i)$ and β is the background term. As β increases, the sampling will tend to a uniform selection from the image set.

10.3 Implementation Details

Flutter is implemented in C# using Managed DirectX for hardware-accelerated drawing. By careful management of texture memory, this can run efficiently on devices with relatively limited graphics capabilities (as in the case of the current generation of UMPC's). Frame rates of 60Hz are achieved on devices such as the Sahara tablet PC and the Samsung Q1, even with several hundred photographs visible, which is sufficient for smooth, responsive interaction.

11. IMPLEMENTATION HARDWARE

The system has been implemented running on a standard UMPC (Samsung Q1), whose compact widescreen form-factor and touch-screen display make it well suited to a photo frame style display. The basic system is augmented with an inertial sensor pack, an RFID reader and a detachable stand (Figure 7).

11.1 Sensing

The overall motion of the frame is sensed using the SHAKE device (Figure 8), which provides tri-axis accelerometer, tri-axis gyroscopes and tri-axis magnetometers, and communicates over Bluetooth. However, in the photo browsing application, only accelerometer readings are used; these are sampled at 60Hz and fed to the models which determine the introduction of new images. More sophisticated use could be made of these sensors (for example, shaking in different directions to focus on particular orderings of images). However, the relative clumsiness of the device when mounted on the frame makes this rather impractical. The sensor

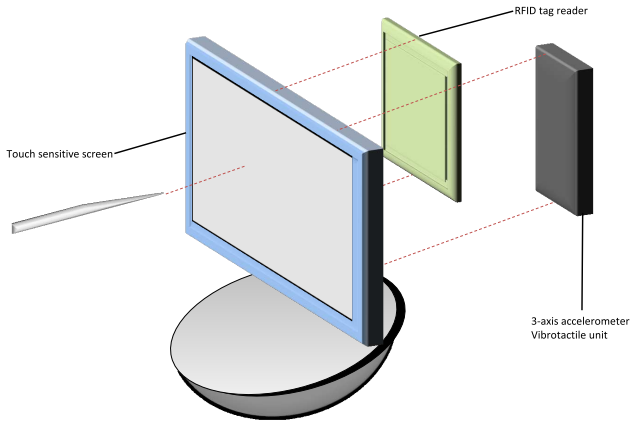

Figure 7: The elements of the photo frame, including the sensing hardware, and the rocking stand for supporting casual interaction.

Figure 8: The SHAKE sensor. This is complete inertial sensing platform running over Bluetooth.

pack also has a built in pager motor which provides the vibrotactile feedback.

The RFID tag reader attached to the rear of the device (a simple USB Phidget) senses whether the device is resting in the stand, which has an embedded tag. As the tag enters and leaves the range of the sensor, the system switches between ambient frame mode, and active contemplation mode.

11.2 Rotating, Rocking Stand

One consideration with a large device such as a tablet PC is the weight of the device. Picking up the frame to interact with it is fine for sitting and contemplating an archive. However, for more passive interaction it can be rather clumsy. Casual interaction with the frame in its "ambient" mode is supported with a weighted stand which is balanced so that it can freely rock and rotate, while maintaining the frame at a comfortable viewing angle. The stand is constructed of a half-sphere, with heavy material at its base, lying in such a way that the hemisphere balances with the top lip at an angle of about 45 degrees to the horizontal when the tablet PC is placed upon it. Gently pushing one side down causes the whole arrangement to slowly rock; this is sensed by the accelerometers and used

to trigger the change of images, once for each "swing" of the stand.

12. CONCLUSIONS AND SUMMARY

Much of conventional HCI is aimed at aiding people in performing tasks or satisfying goals. In this work, the object has been instead to provoke emotional response. Users' own image collections are used as stimuli, and a selection and interaction process is employed that plays the tension between randomness and order and balances user control against the element of surprise. Combining structure that is recorded automatically with photographs with that which is learned from the interactions that people have with the device leads to a system that adapts gradually over time and accumulates a richness of structure. Interactions are simple and direct with just enough depth to support meaningful activity. The tangible nature of the interactions, with rich multimodal feedback and physically modeled animation, makes for an immediate and compelling experience.

The design principles have been derived from an examination of photograph-related activity, and subsequently formulated into a set of design concepts; from these one particular design has been refined and implemented. The result is a complete operational system which fulfills the original objectives. The design presented here, although feature-rich, explores only a subset of the possibilities following from the design objectives; many enhancements could be made, such as support for multi-display, multi-user interaction or more sophisticated use of the inertial sensing capabilities.

Acknowledgments

This work was undertaken while John Williamson was an intern at Microsoft Research. John Williamson is also grateful for support from EPSRC grant "Multimodal, Negotiated Interaction in Mobile Scenarios." The authors would like to thank Dominic Robson for his work in the sound design of the interface. The authors would also like to thank the members of the SDS group at Microsoft Research, especially Richard Harper, Abigail Sellen, Richard Banks, Abigail Durrant and Alex Taylor, who all provided useful input on the design of the system.

13. REFERENCES

[1] Anand Agarawala and Ravin Balakrishnan. Keepin it real: Pushing the desktop metaphor with physics, piles and the pen. In *CHI 2006*, pages 1283–1292, 2006.

[2] Marcia J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5):407–424, 1989.

[3] Frank Bentley and Crysta Metcalf. Flexible views: Annotating and finding context-tagged mobile content. Technical report, Motorola Labs Application Research Center, 2006.

[4] Matthew Cooper, Jonathan Foote, Andreas Girgensohn, and Lynn Wilcox. Temporal event clustering for digital photo collections. *ACM Trans. Multimedia Comput. Commun. Appl.*, 1(3):269–288, 2005.

[5] Steven M. Drucker, Curtis Wong, Asta Roseway, Steven Glenner, and Steven D. De Mar. Mediabrowser: reclaiming the shoebox. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, pages 433–436, 2004.

[6] David Frohlich, Allan Kuchinsky, Celine Pering, Abbe Don, and Steven Ariss. Requirements for photoware. In *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 166–175, 2002.

[7] A. Girgensohn, J. Adcock, M. Cooper, J. Foote, and L. Wilcox. Simplifying the management of large photo collections. In *INTERACT '03*, pages 196–203, 2003.

[8] Karen D Grant, A. Graham, T. Nguyen, A. Paepcke, and Winograd T. Beyond the shoe box: Foundations for flexibly organizing photographs on a computer. Technical Report CSTR-2003-05, Stanford University, 2003.

[9] Cathal Gurrin, Gareth J. F. Jones, Hyowon Lee, Neil O'Hare, Alan F. Smeaton, and Noel Murphy. Mobile access to personal digital photograph archives. In *MobileHCI '05: Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, pages 311–314, 2005.

[10] David Kirk, Abigail Sellen, Carsten Rother, and Ken Wood. Understanding photowork. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 761–770, 2006.

[11] A. Kudhinsky, C. Pering, M. Creech, D Freeze, B. Serra, and J. Gvvizdka. Fofofile: A consumer multimedia organization and retrieval system. In *CHI'99*, 1999.

[12] Baback Moghaddam, Qi Tian, Neal Lesh, Chia Shen, and Thomas S. Huang. Visualization and user-modeling for browsing personal photo libraries. *International Journal of Computer Vision*, 56(1/2):109–130, 2004.

[13] Mor Naaman, Susumu Harada, QianYing Wang, Hector Garcia-Molina, and Andreas Paepcke. Context data in geo-referenced digital photo collections. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 196–203, 2004.

[14] Mor Naaman, Andreas Paepcke, and Hector Garcia-Molina. From where to what: Metadata sharing for digital photographs with geographic coordinates. In *International Conference on Cooperative Information Systems*, 2003.

[15] J. C. Platt, M. Czerwinski, and B. A. Field. Phototoc: automatic clustering for browsing personal photographs. In *Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on*, volume 1, pages 6–10 Vol.1, 2003.

[16] Kerry Rodden, Wojciech Basalaj, David Sinclair, and Kenneth R. Wood. Does organisation by similarity assist image browsing? In *CHI*, pages 190–197, 2001.

[17] Kerry Rodden and Kenneth R. Wood. How do people manage their digital photographs? In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 409–416, 2003.

[18] Yanfeng Sun, Hongjiang Zhang, Lei Zhang, and Mingjing Li. Myphotos: a system for home photo management and processing. In *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, pages 81–82, 2002.

[19] John Williamson, Roderick Murray-Smith, and Stephen Hughes. Shoogle: excitatory multimodal interaction on mobile devices. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 121–124, 2007.

[20] Xing Xie, Hao Liu, Simon Goumaz, and Wei-Ying Ma. Learning user interest for image browsing on small-form-factor devices. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 671–680, 2005.