Leveraging Community Annotations for Image Adaptation to Small Presentation Formats

Patrick Schmitz

Yahoo! Research Berkeley
1950 University Ave. Suite 200
Berkeley, CA 94704

pschmitz@yahoo-inc.com

UC Berkeley School of Information
102 South Hall
Berkeley, CA 94720

pschmitz@sims.berkeley.edu

ABSTRACT
In this paper, we describe an investigation into the use of community annotations of photos for creating zoom and pan animations that present a time linear view of the regions of interest. We describe the use of the notes feature of the Flickr photo sharing service, and describe the FlickrBrns system that leverages the metadata in Flickr to generate animation specifications for image adaptation, and provides a web viewer and authoring controls for more precise specifications.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *Animations* .
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Evaluation/methodology.*

Keywords
Image adaptation, community annotation, Flickr, FlickrBrns.

1. INTRODUCTION
Photographic images are static, but many contain narrative features or qualities that can be expressed over time. These may be specific regions of interest such as faces or objects, where zooming in on each region and panning among the zoomed regions provides a time-linear adaptation of the entire photo content. The regions may also be more abstract, representing the narrative path that the eye follows in understanding a photo.

In addition to the inherent interest of presenting and exploring photos in this manner, the technique can be used to adapt photos for display on small format devices and other constrained layout use-cases. Mobile phones and other handheld devices are in very wide use, and are a common platform for information presentation in a broad range of applications. The small form factor requires some adaptation of certain content, especially high-resolution images. While some images may lend themselves to resizing, many others do not, especially on very small devices. We believe that author-defined notes can be leveraged for this adaptation.

1.1 Related Work
Several systems [5,8] describe mechanisms to automatically detect regions of interest, and then leverage these to produce an end user experience similar to what we describe. Another [3] uses face detection algorithms to assist a human author in describing possible regions of interest. Many such systems employ some variant of "Rapid Serial Visual Presentation" [1] to linearize space as time. The system in [4] combined face recognition and contrast models to infer possible areas of focus and then produced a video by panning and zooming among focus areas.

These models labor under the assumption that it is generally too costly for humans to annotate the images of interest. However, just as communities of users annotate media and documents to produce shared repositories, so too can these communities annotate regions of photos to indicate both the area of interest as well as a text annotation. Examples of community annotation systems include Flickr for images (www.flickr.com), YouTube for video (http://www.youtube.com/), del.icio.us and Yahoo! MyWeb 2.0 for shared web links, and many others. These systems have gathered millions of documents and the respective communities have generated rich annotations (primarily in the form of tags, but also as comments, regions, etc.). The rapid growth of these and similar systems clearly demonstrates the viability and scale of community annotation.

Of particular interest is the *Notes* feature of the Flickr service, whereby users (either the content creator or others viewing a photo) define a region and an optional text caption. In addition to the content of these annotations, the action of specifying regions is an implicit marker for a photo that is well suited to a zoom and pan adaptation technique for small format browsing contexts.

In this project we explored the use of notes on the Flickr service, and characterized classes of notes annotations. We then developed the FlickrBrns[1] service to automatically generate zoom and pan specifications from the Flickr metadata that can be adapted to a wide variety of viewing platforms and applications. A web-based demonstration of FlickrBrns uses XSLT to convert the abstract specification into XHTML+SMIL [7], viewable in MSIE [6]. Finally, an authoring tool was developed that supported finer level control over the animation parameters. The prototype is available at: http://dream.sims.berkeley.edu/~pschmitz/FlickrBrns/.

The next sections describe the exploration of notes usage, the FlickrBrns system, some observations and plans for future work.

---

[1] Documentary filmmaker Ken Burns popularized the technique.

2. NOTES USAGE ON FLICKR

We built a simple application that queries the Flickr system using the public APIs[2], and collects photos that had notes added. We did not attempt to be exhaustive, but collected about 850 photos with notes and then evaluated these. The notes regions were normalized to unit dimensions to simplify presentation and review, and cached in a local database for performance reasons. A simple web-based tool presented a gallery of the collected images with the notes regions indicated, and supported an evaluation interface to specify the *utility* of the regions for the animation technique, based upon region definition, size, etc. An excerpted screen shot of this tool is presented in Figure 1.

Figure 1. Evaluation gallery for Flickr Notes usage.

2.1 Results of evaluation

Depending upon personal settings, notes may be added only by the image creator, or by anyone. There is no clear guidance on the use of notes and the only way to see them (up to now) is on the Flickr site, and over the default-sized image. This results in several quite different genres of notes usage that vary along two dimensions: authorship and intent.

Notes that are added by the author tend to reflect two general modes of use:

1. Notes that indicate interesting regions of the image, with a comment on the significance. These lend themselves most directly to the animation technique. We think a service to leverage these notes for image adaptation to phones would further motivate this genre, and increase usage overall.

2. Notes that simply indicate a general property of the photo that the author wanted to announce to all viewers. These include copyright notices, an indicator of the depicted location, and other comments that appear almost like watermarks. The regions defined for these notes tend to have no significance in the actual photo.

Notes that are added by other users follow a somewhat similar pattern:

1. Others also defined notes for proper regions, although there is some variation in the *quality* (i.e., how well the note defines the actual region of interest). This may be a function of the current UI model in which the region is never seen in isolation.

2. Other reviewer/community defined notes were not as closely associated with a region of the image, but rather with a kind of shared discussion among annotators and the photo creator. In some cases, a set of viewers have added a series of humorous annotations to an image in which the regions had limited significance.

In the limited subset of Flickr photos (with notes) that we evaluated, roughly 6 percent were considered to produce interesting and artistically pleasing animations, nearly 60 percent would be useful to some degree as a means of adaptation to a small format presentation, and just over 34 percent did not use notes in a meaningful way for this technique.

That nearly two-thirds of the notes (among this subset of the entire collection) are useful is significant in itself. Of the remaining cases, it may be possible to use other metadata such as the notes authors, the notes text and the size and distribution of the notes regions to filter out some of the less useful cases and improve the effective percentage of usable notes.

The results with FlickrBrns are promising enough that a more complete evaluation of the broader collection of Flickr images is merited, including a sociological study of notes usage.

3. THE FLICKRBRNS SYSTEM

We designed a system that includes a database to cache a set of images that have notes, a server layer that generates an XML schema describing the animation moves, and an XSLT transform that converts the XML schema to a presentation language. We provided the intermediate abstraction so that we could more easily target different presentation platforms. We built a prototype presentation layer using the animation features of XHTML+SMIL supported in MS Internet Explorer versions 5.5+. The resulting animation produces a 'movie' that emulates the effect used by many documentary film producers and in particular by Ken Burns.

Users choose from among their Flickr photos that have notes, and then use a simple UI to set basic parameters like the pacing of animation and the order in which to present the notes (or they can accept the default values for these and use the system in a fully automatic manner).

We were interested in both the aesthetic aspects of the FlickrBrns animations, as well as in other applications such as image adaptation for small format devices. In addition to purely practical use-cases, we were curious whether more artistic animations might be created as well, and so we created a simple but flexible user interface for finer control over parameters.

3.1 FlickrBrns service

A PHP based service accepts parameters to describe aspects of the desired animation, and uses these and the metadata for the image to generate an abstract description of the animation to be performed. The parameters specify how the notes/regions should be ordered in the animation, how quickly to animate the transition (i.e., the zoom and pan) from one to the next, and whether to pause after each note. In addition, the viewport dimensions are specified, so that the resulting animation fits a given layout context.

Many parameter values are specified logically, although some flexibility is supported so that more artistic animations are also possible with the system. For example, the notes order is generally specified logically as left-to-right, top-to-bottom, etc. However it can also be specified as a sequence of notes index values, allowing arbitrary animation ordering. An author would use the creative interface to specify parameters, where an image browsing service for phones would just specify viewport constraints, etc.

3.1.1 Default parameter values

Default values are assumed for all parameters other than the image id. Thus, some reasonable animation can be automatically produced for any image with notes. For a given platform such as mobile phones, the service would be adjusted so that some parameters would use defaults that reflect the platform capabilities (especially the viewport size). However, we are exploring more intelligent default values based upon the available metadata.

For example, in the initial implementation the pause duration is specified as a single value for the entire animation (for all steps). While this works for most cases, the pause duration should be some function of the notes *text* that is presented for each region. When notes provide detailed context for each region, they often vary considerably in length. We are currently exploring several methods of defining the pause duration to accommodate this.

As another example, the default notes order is left-to-right. A more intelligent mechanism should consider the layout and sizes of the notes regions, determine whether clear lines, arcs or paths exist in the set of regions, and use this as a basis for ordering.

3.1.2 XML Animation schema

The XML produced by the service is straightforward. The animation for an image is specified as a series of steps. There is one step for each note/region, with one extra step each if the full image is shown at the beginning or end. These animations can be grouped together into a series, so that one animation file describes a sequence of images with animations.

```
<imgset loop="true">
<viewport w="160px" h="120px" />
 <img>
  <imgURL>http://static.flickr.com/…</imgURL>
  <pgBaseURL>http://www.flickr.com/…</pgBaseURL>
  <steps>
   <step x="0px" y="0px" w="160" h="120px"
         holdDur="1" />
   <step x="-11.83px" y="-8.87px"
         w="171.83px" h="128.87px"
         animDur="3" holdDur="1"
         accelerate="0.1" decelerate="0.1" >
     The Momma.
   </step>
   <step x="-123.4px" y="-98.2px"
         w="400px" h="300px"
         animDur="3" holdDur="1"
         accelerate="0.1" decelerate="0.1" >
     The young one.
   </step>
   …
  </steps>
 </img>
</imgset>
```

Figure 2. Example excerpt of animation specification XML

Each step describes the target size and location relative to the clipping viewport. If a non-zero animation duration is specified, a smooth animation is performed from the previous step to the current values; otherwise the step is to happen immediately, and so the new values replace the previous ones. Additional parameters support ease-in/ease-out to produce smooth animations, pauses at each step, and the ability to loop the image sequence. Figure 2 provides a simple example of the XML animation specification. The first step specifies the entire image scaled to fit the viewport, followed by steps that zoom in to regions.

3.2 Animation implementation

The presentation implementation uses XSLT to convert the animation schema to XHTML+SMIL for presentation in MSIE. The animation functionality breaks down as follows:

Zoom and Pan control – Each photo is wrapped in a `<div>` element with relative positioning to define a viewport origin and provide clipping. Individual animations are then applied to each of the CSS `top`, `left`, `width` and `height` properties for the image. The transform animations all use additive "to" animation, so that each can be defined independent of the others and simply build upon (compose with) the previous animations. For smoothing, we experimented with various settings of `t:accelerate` and `t:decelerate` and found that in general, small values produced enough smoothing to be pleasing without dramatically altering the character of the animation.

Introducing pauses at each note - We support these with empty timed elements that just extend the parent sequence for the desired duration (a simple but effective technique).
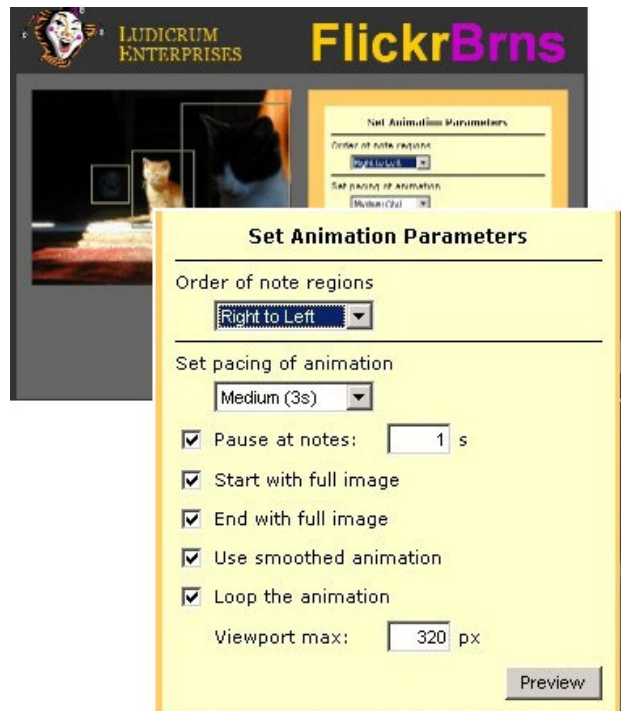

Figure 3. Fine authoring control user interface

Synchronization and layout of the notes text - In the first prototype, we simply made the notes appear for the duration of the pause. However, pauses may be very short or be omitted altogether. We detect these cases and make the note text appear for some portion of the animation leading to the associated region, and through any pause.

3.3 Authoring control

Users start from a simple page that asks for their Flickr id, and then presents them with the set of their own photos that have notes defined for them. They can then open these in a simple form that supports more precise definition of the parameters as described above. We reworked this UI design several times as we tried to find the right balance between simplicity and flexibility. The resulting UI is illustrated in Figure 3.

4. OBSERVATIONS AND FUTURE WORK

Some of our initial assumptions and models evolved as we gathered experience with the use of notes among Flickr users. Issues with smarter default parameters and the need to pursue a more in-depth analysis of the Flickr collection were mentioned above. Other issues from our ongoing analysis are described here.

4.1 Viewport issues

We currently allow specification of the viewport size only as a maximum dimension, as we assumed that it should match the aspect ratio of the full image. We had also experimented with and then removed viewport *animations* (changing the size and shape of the clipping viewport) to improve performance and simplify some of the code. However, for many cases of notes in the Flickr collection, we found that the aspect ratio of some notes was quite different from the aspect ratio of the viewport (i.e., from the aspect ratio of the full image). This means that the animation does not really emphasize the note region very well.

For contexts with more flexible layout constraints, we will experiment with viewport animation so that it is possible to highlight regions that have very different aspect ratios from the full image. We are also considering static viewports different in aspect ratio from the image, to support cases like panning a vertical stripe across a panoramic or landscape-mode photo.

4.2 Image size

Most simple image renderers do a much better job of filtering an image to a smaller than original size, than to a larger size. As a result, if the original is not large enough to be at or below full size when zoomed in on a (relatively small) note region, the resulting animation can pixelate badly. At the same time, when the image is very large and easily supports considerable zooming, it is more expensive to manipulate it in real time, and so the animation becomes much slower or not as smooth.

We have considered several approaches to alleviate these issues, and are experimenting with a threshold on the maximum allowed zoom, with a highlight displayed for the note region when the threshold precludes zooming fully into a region.

4.3 Support for other platforms

The animation support in MSIE works well for small regions of animation and nicely integrates with HTML pages. However, for larger animations, and as well for other platforms, we are exploring a simple Flash application that interprets the XML schema. Finally, we are evaluating phone and PDA platforms that can reasonably support the animations.

5. CONCLUSIONS

In many cases, the notes defined by Flickr users produced pleasing or useful animations over the significant regions of interest. Certain classes worked particularly well in fully automatic mode, including group portraits (e.g., of sports teams or families), and many photos with highlighted details make for interesting movies on a small viewport as for a phone or PDA.

It was somewhat more difficult to produce truly artistic movies from the notes. In particular for some nature shots, authors wanted to make the FlickrBrns movie recapitulate the line an eye would follow through an image. To get this effect, users began to define notes that had little or no utility when viewing the still image. We had not anticipated the extent to which this was the case.

Nevertheless, for both utilitarian as well as aesthetic use-cases, author and community-defined notes are a valuable source of information about regions of interest, and provide an alternative to purely analytical approaches to image summary and adaptation.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Bruijn, O., and Spence, R. 2000. Rapid serial visual presentation: a space-time trade-off in information presentation, In *Proceedings of Advanced Visual Interfaces* 2000, pp189-192.

[2] Flickr Services API, documented at www.flickr.com/services/api.

[3] Girgensohn, A., Adcock, J., and Wilcox, L. 2004. Leveraging face recognition technology to find and organize photos. In *Proceedings of the 6th ACM SIGMM international Workshop on Multimedia Information Retrieval* (New York, NY, Oct 15-16, 2004), 99-106.

[4] Hua, X., Lu, L., and Zhang, H. 2003. Photo2Video. In *Proceedings of the Eleventh ACM International Conference on Multimedia* (Berkeley, CA, Nov. 2-8, 2003), 592-593.

[5] Liu, H., Xie, X., Ma, W., and Zhang, H. 2003. Automatic browsing of large pictures on mobile devices. In *Proceedings of the Eleventh ACM International Conference on Multimedia* (Berkeley, CA, Nov. 2-8, 2003), 148-155.

[6] Microsoft Corp. XHTML+SMIL support documented at: msdn.microsoft.com/workshop/author/behaviors/time.asp

[7] Newman, D., Patterson, A., Schmitz, P. 2002 XHTML+SMIL Profile W3C Note. Available at: www.w3.org/TR/XHTMLplusSMIL/.

[8] Xie, X., Fan, X., Ma, W., Zhou, H. 2003. Adapting images on proxies for small form factor devices. In *Proceedings of the 2003 Joint Conference of the Fourth International Conference on Publication*, Dec. 15-18, 2003, V1, 428-432