Generating Video from Digital Photographs

Jonathan Foote, Matthew Cooper, John Adcock, and David Hilbert
FX Palo Alto Laboratory
3400 Hillview Ave. Bldg. 4
Palo Alto, CA USA
{lastname}@fxpal.com

Keywords
audio analysis, image analysis, automatic video generation

ABSTRACT
We present methods for automatically turning digital photographs into a compelling video animation. Images are animated using dynamic effects like cropping, panning, zooming, rotation, and warping, synchronized with a user-selected soundtrack. Appropriate effects are automatically determined from features extracted from the images. For example, images with detected faces or strong symmetry are good candidates for a zoom effect. The system architecture flexibly supports other animation criteria such as pan direction from spatial frequency analysis. These dynamic effects are automatically synchronized to an audio soundtrack, resulting in a compelling video. We also consider lightweight output types such as Macromedia SWF and SMIL as well as more conventional video formats.

1. INTRODUCTION
The widespread popularity of digital cameras has resulted in a large number of personal image collections. At the same time, digital technology offers new ways of organizing and sharing those photos. In a recent study [1], HP researches ask "What do users want to do differently with photos once they have captured them in the digital realm?" As a partial answer, they observe that "Future technology should help users. . . to extend the sharability of digital photos across a range of use contexts." We look at one such extension in this paper, which presents a way to automatically render static digital photos into a dynamic audio/video presentation.

We present novel methods for animating still images using dynamic effects, selected automatically to enhance the characteristics of each image. Ideally, we wish to automate techniques used by documentary filmmakers to enliven still images. Given a static image such as an archival photograph, documentarians can creatively pan and zoom across the image, resulting in a dynamic presentation for film or television. For example, given an image with three human figures, a documentary "pan-and-scan" video shot might start with a close-up of the leftmost figure's feet, then pan towards the face, then pan from left to right to show each figure's face in turn, then zoom out to include all figures against the background scenery. Face detection technology that can automatically locate human faces in imagery ([7, 15]) allows this to be done without human intervention. Documentarians commonly use music to enhance the emotional tone of the final video, which is something we support as well. In the above example, the foot-to-face pan could be synchronized with a musical crescendo to emphasize a zoom in on the final face.

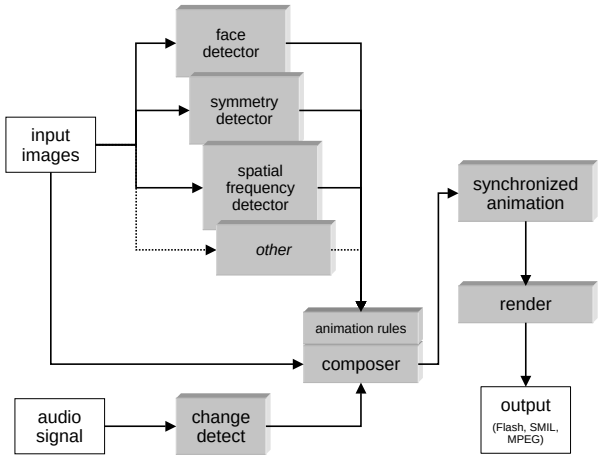

Figure 1: Block diagram of automatic animation system.

While many photo management applications now include a "slideshow" output, typically these are nothing more than a sequence of still images presented at regular intervals. By including synchronized music and dynamic effects, we hope to automatically create more compelling content without the laborious effort of manually animating and synchronizing images. In developing this system, we make two key assumptions. The first is that a good soundtrack can enhance the emotional impact of images. This fact is a truism in the film industry, and has been backed up with a number of studies. There is evidence that good audio can enhance

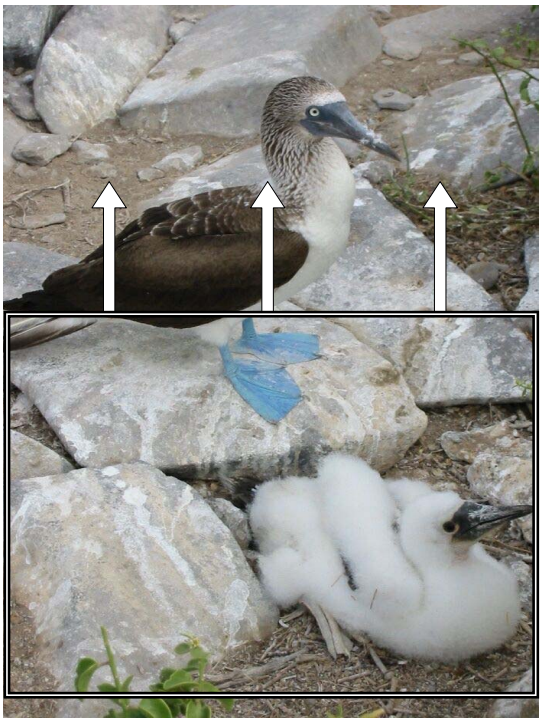

Figure 2: Animating still images by panning across cropped regions.

the appearance of images. One study at MIT claims to have shown that listeners judge the identical video image to be higher quality when accompanied by higher-fidelity audio [2]. Our second assumption is that synchronizing picture changes with audio events enhances the perception of both. Not only is this a common practice of cinematic sound editors worldwide, but it is also backed by empirical evidence. For example, Lipscomb [3, 4] presents user studies demonstrating that the "effectiveness" of a film clip is enhanced when audio and video events are synchronized. Finally, the rate of the resulting animation is determined primarily by the selected soundtrack music, so the user can control the mood and tempo of the result by selecting the appropriate soundtrack.

2. RELATED WORK

There is no shortage of image presentation products such as Microsoft PowerPoint or Kai's Power Show [13], as well as various video editing tools that can turn still images into video. While most allow images to be scheduled and presented at a specified time, only a few allow the inclusion of audio. Those that do fall into three major classes. The first allows a user to manually attach an audio annotation to an individual presentation image. The second allows the user to select a single background digital audio file. Each image then appears for a uniform duration, typically the audio

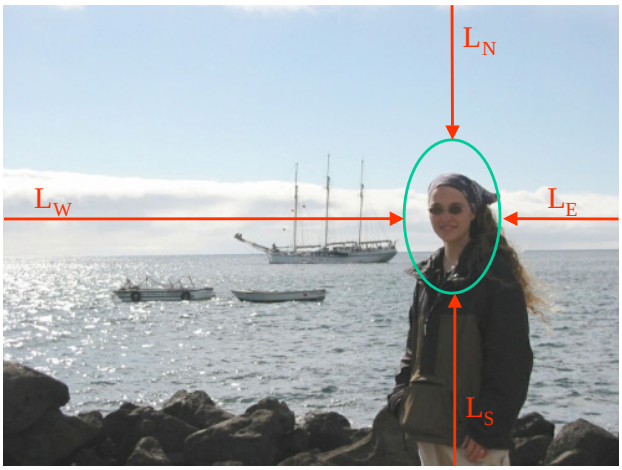

Figure 3: Determining the best pan direction from face location.

duration divided by the number of images (e.g. [13]). Finally, several packages allow the user to manually schedule the duration that each image will appear in a slideshow with a soundtrack (e.g. [14]).

Pan-and-scan effects are supported by professional video packages such as Adobe After Effects or Final Cut Pro. Here, the desired pan path or zoom effect must be fully specified by the user. Many report this to be a tedious and time-consuming process not likely to be pursued by casual users. Apple's iMovie has a "Ken Burns Effect" for still photos that zooms in to the center of the image for a user-specified duration. Specifying the zoom ratio and duration, as well as synchronizing this to any audio, are operations left to the user. Furthermore the center of the zoom can't be changed, which is suboptimal for interesting objects not centered in the picture.

The precursor to this work is our automatic video creation system, where video shots are automatically synchronized with musical changes [10]. The image content of the video was not really considered except to find the time location of shot boundaries, and to exclude shots with excessive motion or poor exposure. In contrast, the system presented here depends on different kinds of image analysis to infer the animation effect most appropriate to each image. We do, however, use the same approach to audio analysis, summarized in Section 4. There exist several automatic video editing systems available commercially, for example Muvee.com and ACD VideoMagic. The literature for these products claim automatic analysis of video and audio, but lack sufficient technical detail that we can make a meaningful comparison with our approach.

3. OVERVIEW

The system starts with a selected audio soundtrack and a number of still images. Typically the audio is a musical work and the images are digital photos. The system detects significant changes in the audio and then schedules images to be displayed synchronously with them. Alternatively, the tempo of the music can be determined using any number

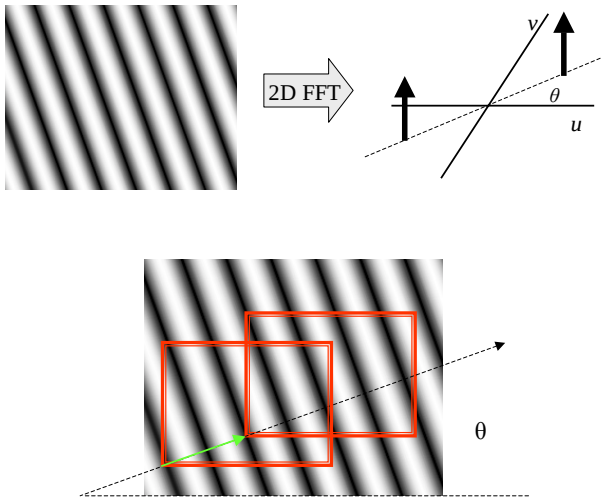

Figure 4: Panning in the direction of maximum spatial frequency.

of techniques, and images scheduled to change at synchronous rate. Currently, our system assumes that images come with an implicit order, either user-selected or by timestamp. Images are displayed in that order (though we discuss reordering images to better fit the soundtrack in Section 8). Furthermore, we assume that each image has the appropriate "portrait" or "landscape" orientation and do not need to be rotated (though automatic methods to determine correct orientation have been reported [5]).

During the time an image is displayed, pan and zoom effects add motion and interest to the image. Figure 1 shows a block diagram of our animation system. The system includes several image analysis modules, each tuned to detect particular image features that assist animation. The face detection module detects the location of faces in the image, as these are often the focus of interest. Panning to or zooming in on a face can dramatically emphasize the people in a photograph. Another detector is specialized to detect vertical-axis symmetry. If an image has a large degree of vertical symmetry, it is often a good candidate for zoom operations (into or out of the center of symmetry). Examples include doorways or arches, frontal views of vehicles, and streets or paths viewed axially. A further detector finds the direction and magnitude of spatial frequency components in the image. A direction with large magnitude of spatial frequency is a good candidate for panning operations, as the resulting video will have interesting time variation. For example, a number of standing people (or trees, or windows) photographed against a wall will have the highest magnitude component in the horizontal direction. Panning in this direction will thus produce more interesting video because it will have more time variation.

In operation, each detector has an output proportional to the strength of the feature it is designed to detect. Each detector analyzes every input image. For each image, the composer module selects an animation type based on the relative strengths of the various detector outputs, and ad-hoc rules. The simplest set of rules assign a predefined effect based on the strongest detector output. For example, if the face detector reports a face with high confidence, and little symmetry or spatial frequency components are detected, then a zoom effect is chosen (into the detected facial region). In the absence of strong outputs from any detector, an effect is chosen randomly. Detector outputs can be weighted to preferentially encourage particular effects. More elaborate rules might vary the zoom direction or effect type randomly, programatically, and/or based on the available time or effects applied earlier . Note that this architecture easily supports additional detector modules if other or more sophisticated analyses are available.

4. AUDIO ANALYSIS

To detect audio changes, any convenient method may be used, for example the amplitude envelope may be analyzed to detect onsets of high-energy regions as good candidates for change points. In the current system, the similarity-based methods of [8] and [9] are used to detect significant audio changes and repeated segments in the soundtrack audio. The result is a time-indexed measure of audio change versus time. Peaks in this "novelty score" represent audio changes, and may be ranked by amplitude. Thus if N pictures are to be animated, the $N-1$ peaks with the highest amplitude can be used to sequence the pictures. In actual use, other heuristics are used to select the audio changes, for example changes closer than some minimum time are discarded so that pictures are not displayed too rapidly. Recent work has also shown that repeated audio segments (such as verses and choruses in popular music) can be detected with reasonable accuracy [9]. The boundaries between song segments are excellent transition points for an image change or other effect.

5. AUTOMATIC IMAGE PANNING AND ANIMATION

It is not necessary to display a new picture at each detected audio change. For example, each detected facial region can be considered an image in its own right. A simple heuristic is to crop the image into regions containing a face, and display each new face with a corresponding audio change. For a smoother presentation, the image can be "panned" over time using linear interpolation to move smoothly between the face regions. Other transformations can be used to animate a single image, for example, rotation or zooming. In this section, we present our image analyses in more detail, and show how they are used to create effective animations.

5.1 Simple Panning

Analogously to a video camera moving across a large scene, a still image can be "panned" by continuously moving a cropped region. This introduces motion that enlivens a still image to a surprising degree. A simple method is to use the discrepancy between the aspect ratios of photographs and video. Typical photographs have a 4×5 aspect ratio, while video typically has a 3×4 ratio. Thus an image must be cropped to fit the video, severely so if it is in the "portrait" orientation, as in Figure 2. The cropped portion, normally discarded, can be used to animate the image by sliding the view window so that the cropped portion comes into view. This has two advantages: the animated image is in motion and thus more compelling, and it is also used more com-

Figure 5: (L) "Hammocks" image, and (R) radial frequency histogram of "Hammocks" image.

pletely, as the cropped portion becomes visible. The speed of the pan can be determined by the corresponding audio segment such that longer segments result in slower pans. This is done by finding the pan (or zoom) rate that results in the desired pan (zoom) distance at the length of the segment. This kind of animation works for all images, and is thus particularly useful when the automatic analysis detects no outstanding features such as faces or symmetry.

5.2 Face Detection

Image analysis can result in more sophisticated pan and zoom effects. Face detection algorithms can detect the location of faces in an image; likewise the user can indicate faces or other interesting "hot spots" in an image (such as a mountain in a landscape or a boat on the water). Once these spots are determined, a reasonable pan approach is to find the longest possible pan from the picture edge to the hot spot. For example, if the subject is in the far right of the picture, then the appropriate pan would be from the left. This takes advantage of the photographer's tendency to compose interesting features in the image: if there is a subject on the right, there's likely photogenic scenery or another subject on the left. Thus a good heuristic to generate pan effects is to compute the maximum distance of the hot spot to all image edges. Figure 3 shows an example. The hot spot is indicated by the ellipse, which is the result of either a face location algorithm or a user-provided label. Examining the distances from each side, the algorithm determines that L_W, the distance from the west, is the maximum and thus the best pan is from the left. If none of the distances are significantly larger, then the subject is considered central and a zoom can be used instead of a pan. In this case, the picture is zoomed in or out from the hot spot over the length of the segment. For the examples here and in our video figures, we used the probabilistic face detector developed by Henry Schneiderman as part of his graduate thesis at CMU [15]. We have been impressed by the consistent accuracy of his system.

5.3 Spatial Frequency Analysis

Other image analysis can be used for automatic panning. One method is to look at the angular distribution of spatial frequencies. The 2-D Fourier transform gives not only the frequency components of an image, but their spatial direction as well, as shown schematically in Figure 4. Here, a sinusoidal image component produces two strong peaks in the frequency plane. The angle of this spatial frequency component can be determined from the angle θ of the frequency-domain peaks. Panning in this direction will produce movement in the direction of the major spatial frequency component. In practice, we compute a "direction marginal" by summing the magnitudes of the 2-D Fourier transform radially from the origin. To reduce edge artifacts, we taper the image with a radial Gaussian function that decreases towards zero at the image edges. We also zero-pad the image before computing the frequency components using the 2-D FFT. More sophisticated processing might address the fact that components at extremes of spatial frequency are not as significant. Important mid-frequency components can be emphasized by multiplying the frequency-domain results with an annular (ring-shaped) window, equivalent to a band-pass spatial filter.

Once the spatial frequency components are computed from the DFT, they are thresholded to remove noise. An angular histogram is computed by summing the frequency components in wedge-shaped bins around the origin, discarding low-frequency components. Polar plots of these histograms are shown in the right-hand panels of Figures 5 and 6. In particular, the "Pyramids" image clearly show how edges give rise to strong frequency components. In the radial frequency plot, the diagonal peaks are mainly due to the sloping pyramid edges, while the vertical component is caused by the ground-sky edge. (In our system, the maximal peak at 45° would result in a diagonal pan across the image, from bottom left to top right through the center of the image.)

Typically, edges are the image features that yield large spatial frequency components; also stripes and bands, like the hammocks in Figure 5. The direction with the largest com-

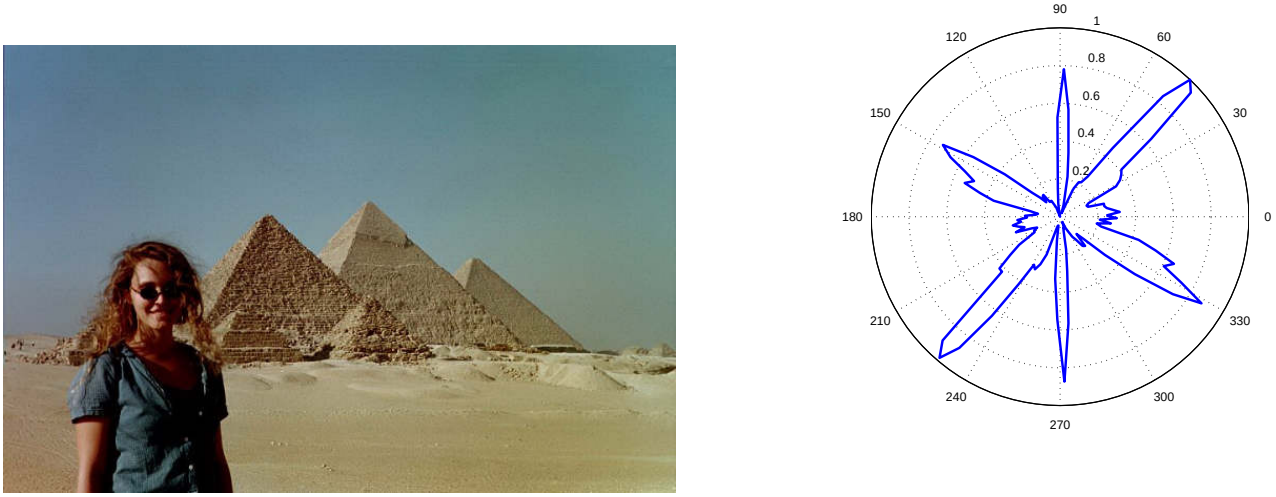

Figure 6: (L) "Pyramids" image, and (R) radial frequency histogram of "Pyramids" image.

ponent sum is chosen as the direction to pan: this ensures a pan with the highest visual change, and thus presumably visual interest. For example, given a photograph with horizontal components of beach, ocean, and sky, the largest components will be in the vertical direction ($\theta \approx \pm 90°$). Thus panning vertically will show more visual change (beach, ocean, sky) than a horizontal pan ($\theta \approx 0°, 180°$). In general, this method results in pans that are perpendicular to major edges or stripes. Though we find this assumption typically does the right thing, there are cases (such as people sitting on a beach or standing on a mountain ridge) where panning orthogonally to the major edge does not result in an optimal pan; this is one reason we have given the facial detector priority over the frequency analysis.

5.4 Symmetry Detection

Image symmetry gives another way to locate regions of interest. Many approaches to symmetry detection can be found in the literature, for example [16, 17, 18]. These can be used to find the axis of local symmetry, which are natural locations to zoom into or out from, as shown in Figure 8. Amateur and professional photos exhibit symmetry surprisingly often, for example in head-on shots of faces, vehicles, and animals, many architectural features such as doors, windows and halls, natural features such as sunsets and trees, as well as landscape features such as buildings and bridges. A particular case are streets, roads, rivers, and paths when photographed axially – that is, in the same direction as the object runs. In these examples, a zoom into the center of symmetry has the visual effect of moving along the path or street.

Our symmetry detection approach uses straightforward correlation to find both the strength and the center of vertical-axis symmetry. Each image is zero-padded horizontally and correlated with its vertical-axis reflection at different horizontal shifts. The result is a function of the shift, and will have a maximum at the center of symmetry (if care is taken with indexing). Figure 7 shows the axis of symmetry detected for a street photograph, while the left image of Figure 8 shows a similar result for a photograph of a scene in Canberra, Australia. Just as usefully, the maximum correlation value measures the degree of symmetry. This value must exceed a substantial threshold before the image is considered symmetric. Though we only consider vertical-axis symmetry here, this technique can be easily extended to find bilateral symmetry at other angles. Other methods can be used to find interesting zoom areas such as spatial moments, or other image symmetries like the orientation distribution methods of Sun[17].

5.5 Composing the Animation

The "composer" schedules images and selects effects based on the automatic analysis of the input audio and images. This is done using a rule-based system. We have attempted to encode a set of rules that offer reasonable results, but they are by no means optimal or even necessarily preferable to others. We have explored only a small set of the design space, and it it highly likely that better rules could be chosen, particularly by those with more video production expertise than the authors. Our assumptions and design decisions are typically based on common-sense heuristics. For example, panning and zooming operations require the image to be cropped to some extent. We choose 1/4 of the image size as the maximum cropping (though this may be zoomed in further). That is, images are not cropped by more than half their extent in length or width. (Images will typically have different aspect ratios than the crop region, so for this discussion "1/4 image size" should be understood as the smallest rectangle with the desired aspect ratio that is not smaller than 1/2 the original image's width or height.) Even though most digital images have enough resolution to support much tighter cropping, we wish to respect (and enhance) the photographer's original composition as much as possible.

In the composer, the face detector has priority, because we assume that people are of primary interest in consumer photographs. Thus a detected face will result in an animation that will emphasize that person's image. The facial detector

Figure 7: Detecting axis of vertical symmetry for zoom operations

estimates face locations and sizes, with a confidence score for each face. In the current system, we chose only the face with the highest confidence, so that only one face is selected per image, though there may be many more visible. (This is primarily for simplicity; more sophisticated implementations could certainly pan across and dwell on each detected face.) We choose an animation that will emphasize this face in relation to the rest of the image. If the face is near the center of the photo, we zoom into the facial region. If the face is near the edge of the photo, specifically if the face center is less than 1/3 the image dimension from an edge, then we pan from the furthest edge towards that face.

If the face detector does not report a face region with sufficient confidence, we then examine the spatial frequency and symmetry detector outputs. Symmetry is given the next highest priority, because there are generally fewer photos that exhibit strong symmetry, and we wish to emphasize it when it has been chosen by the photographer. So if the symmetry measure exceeds another threshold value, we generate a zoom into the center of symmetry. This is done by cropping the photo to the largest size possible that puts the center of symmetry at the center of the crop region. If the center of symmetry is closer to an image edge than 1/3 of the image dimension, then the crop region is chosen to be 1/3 the image size. In any case, the crop region is zoomed in until it is the default crop size, or 1/4 of the original image size. For the sake of variation, a "zoom out" can be generated by reversing the procedure above. In practice, we randomly choose 80% of zooms to be "zoom ins" (magnification) and the remainder as zoom outs; it would be interesting to study, and perhaps mimic, the proportion in professionally produced videos [6].

If the image does not exhibit faces or strong symmetry, we then determine whether a directional pan would be appropriate by examining the strength and direction of the spatial frequency components. If the strength of the largest component exceeds another threshold, then we generate a pan in that direction. This is done moving the crop region along the line that passes through the center of the image, from left to right, at the angle of the maximum spatial frequency component. For example, in the "Hammocks" image of Figure 5, the strongest frequency component is about 40° from the horizontal, so the generated pan would move the crop region from bottom left to top right at this angle.

This initial approach to generating animations is relatively fixed and algorithmic. We are currently working on adding much more flexibility, including a "stylesheet" approach to selecting animation effects. For example, user-controlled parameters could set the maximum pan and zoom rate, minimum time intervals, minimum crop size, and other variables to change the style of the generated animation. Thus setting slow pan rates and longer time intervals would produce a more sedate animation, while selecting the opposite would give a livelier, MTV-like feel to the results.

6. ANIMATION OUTPUT

Because motion in these animations is typically simple affine transformations of still images, full video bandwidth is not required. Several popular animation formats, such as SWF [11] and Synchronized Multimedia SMIL [12], support exactly this kind of animation, and require far less in terms of bandwidth (file size) or CPU needed for rendering. If desired, these formats can still be transcoded to more conventional video formats such as MPEG-2.

Significant liberties can be taken with image display, especially using a medium like Macromedia's SWF ("Flash") format that allows images to be tiled, zoomed, cropped, warped, overlapped, moved, and placed at arbitrary locations at arbitrary times [11]. Thus one image display format could be a large central image surrounded by smaller subsidiary images. With a musical change, the central image can be demoted to a smaller image on the periphery and a new image displayed in the center. Images could be warped or moved in a time-varying manner to dance in synchronization with the music. Text or other graphic elements (speech balloons, clip art) can be introduced or moved in response to musical changes, and can overlay displayed images. Interactive controls such as buttons or sliders can be introduced, and scripted to produce time-dependent actions. For example, a "skip" button might skip to the next image in the sequence, or a "repeat" button might restart the animation and the music from the beginning.

Additional digital video output formats such as DVD or the MPEG format family can also be supported. Because these formats are specialized for motion video, they may be less optimal in terms of required storage when compared to the original source still images and audio, however "sprite" features in the MPEG-4 standard support still image anima-

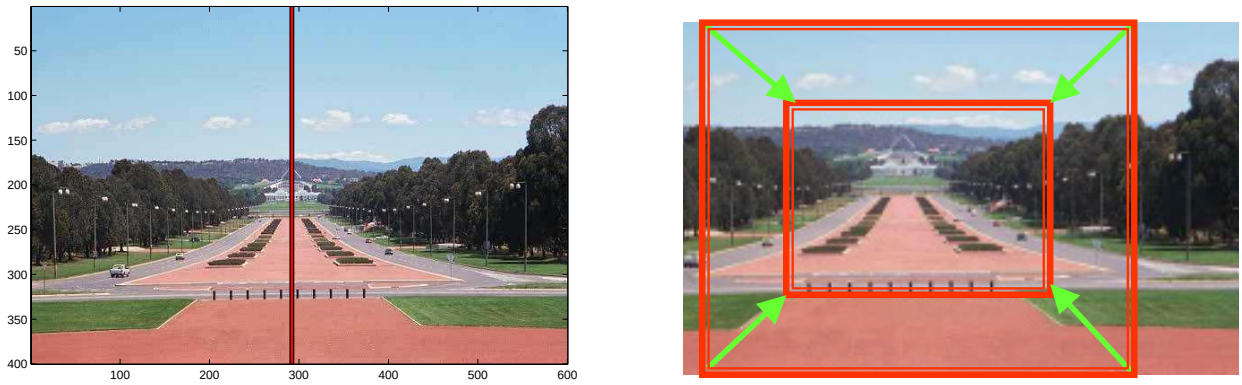

Figure 8: Automatic zooming into the center of symmetry.

tion. DVDs provide an ideal format for slideshow archiving, enabling organized digital photo albums with full audio soundtracks. Slideshows could also be saved to digital video tape or rendered to one of the various streaming video formats for web distribution.

7. RESULTS

We have automatically created a number of animations using a variety of images taken by colleagues and friends, and a mix of soundtrack audio, including popular, jazz, orchestral, and dance music. Because this produces multimedia output, it is not obvious how to present these animated results in a publishable form. Moreover, it would be very desirable to measure how well the algorithms work. However, like our previous work on automatic video creation, these are not objective results which can be benchmarked against some agreed-upon ground truth. Rather, proper evaluation requires subjective results from a significant number of users. Informal judgements show that our method produces convincingly compelling animations. Each animation presents the input photographs with dynamic animations that are noticeably aligned with musical transitions, such as soft/loud or verse/chorus changes in the soundtrack audio. For this paper, we have prepared examples of the various animation effects, and several longer animations. These will be submitted as Video Figures.

8. FURTHER WORK

This work has explored only a small set of a comparatively rich design space. Though we have discussed a purely automatic system, it should be noted that user interaction can be incorporated at any stage. For example, a semi-automatic system might allow users to select particular images for animation, the order of their presentation, or possible effects or enhancements, given an appropriate user interface.

The system discussed here presents images in a pre-defined order. An interesting variation is to sequence images to better match the music. Given a set of digital images, it is possible to analyze each image pair to compute various measures of similarity. Often, the similarity measures are simple measures by which low-level feature vectors may be compared. Alternatively, low level features, including color, brightness, or texture, may themselves be used to select images or determine the image sequence for presentation. As well, available metadata from digital cameras, including timestamps, the identity of the photographer, or GPS data can also be used to cluster or sequence the images. Various clustering techniques based on similarity measures, including spectral clustering as in [9], or hierarchical agglomerative clustering may be used to group images into an automatically determined or user determined number of clusters. From such clusters, various keyframe selection algorithms or feature-based approaches can be used to determine a representative image for inclusion in the slideshow. If desired, this can reduce the inclusion of redundant or overly similar images

Other audio and image analysis can be used, for example detecting crescendos or diminuendos in a segment of the music. This can be done by first segmenting the music by finding significant changes as above. If a segment is longer than some minimum length, the slope of the amplitude or power envelope can be computed using a linear fit. If this is above (below) some threshold, then the segment is classified as a crescendo (diminuendo).

Given that a set of digital images, have been analyzed for image similarity, the images can be ordered so that large audio changes are synchronized with large image changes (transitions between highly dissimilar images). Alternatively, image similarity can be used to either emphasize the distance between each image (for a more dramatic presentation) or to minimize it for a smoother, less jarring result.

Images can be hierarchically clustered into a desired number of clusters to match the song structure of popular music, which may be analyzed using the techniques of [9]. For example, given a number of holiday pictures, daytime beach scenes and nighttime dinner scenes would be naturally separated by brightness and color. These clusters can be matched with the audio structure so that beach scenes are displayed during the chorus and nighttime images are displayed during the verses.

9. REFERENCES

[1] Frohlich DM, Kuchinsky, A., Pering C., Don A. and Ariss S. Requirements for Photoware. *Proceedings of CSCW '02*, New York: ACM Press. 2001 http://doi.acm.org/10.1145/587078.587102

[2] W. R. Neuman, Beyond HDTV: Exploring Subjective

Responses to Very High Definition Television, MIT Media Laboratory Report, July, 1990

[3] S. D. Lipscomb. Perceptual measures of visual and auditory cues in film music. *JASA* **101**(5, ii), p. 3190 http://imr.utsa.edu/∼lipscomb/JASA97/

[4] S. D. Lipscomb and R. A. Kendall. Perceptual judgment of the relationship between musical and visual components in film. *Psychomusicology*, **13**(1):60-98, 1994. http://imr.utsa.edu/∼lipscomb/Thesis/thes00.html

[5] A. Vailaya, H.-J. Zhang, C.-J. Yang, F.-I. Liu, and A. K. Jain Automatic Image Orientation Detection IEEE Transactions on Image Processing, vol. 11, no. 7, pp. 746–755, July, 2002

[6] Y. Matsuo, M. Amano, K. Uehara Mining Video Editing Rules in Video Streams in *Proc. ACM Multimedia 2002*

[7] H. Rowley, S. Baluja, and T. Kanade. Neural Network-Based Face Detection, *IEEE Trans. PAMI* **20**(1):23–38, January 1998

[8] J. Foote. Automatic Audio Segmentation using a Measure of Audio Novelty. *Proc. IEEE Intl. Conf. on Multimedia & Expo*, vol. I, pp. 452-455, 2000.

[9] J. Foote and M. Cooper. Media Segmentation using Self-Similarity Decomposition In *Proc. SPIE Storage and Retrieval for Multimedia databases*, Vol 5021, pp. 167–75, 2003

[10] J. Foote, M. Cooper, and A. Girgensohn, Creating Music Videos using Automatic Media Analysis, in *Proc. ACM Multimedia 2002*

[11] OpenSWF.org SWF Format Specification http://www.openswf.org/spec.html

[12] Synchronized Multimedia Integration Language (SMIL 2.0) W3C Recommendation 07 August 2001 http://www.w3.org/TR/smil20/

[13] Scan Soft Software, Inc. Kai's Powershow http://www.caere.com/products/show/

[14] Apple Computer, Inc. iPhoto http://www.apple.com/iphoto

[15] H. Schneiderman and T. Kanade Face Detection http://www.ri.cmu.edu/projects/project_416.html

[16] H. Zabrodsky, S. Peleg, and D. Avnir, "Symmetry as a Continuous Feature," *IEEE Trans. Pattern Analysis and Machine Intelligence,* Aug 1993, citeseer.nj.nec.com/zabrodsky95symmetry.html

[17] C. Sun. Symmetry Detection Using Gradient Information. *Pattern Recogntion Letters* **16**:987-996, 1995.

[18] C. Sun and D. Si, Fast Reflectional Symmetry Detection Using Orientation Histograms, *Journal of Real-Time Imaging* vol.5, no.1, pp.63–74, February 1999.

[19] R. Gonzales and R. Woods. *Digital Image Processing.* Addison-Wesley, 1992.