The use of a cast to generate person-biased photo-albums

Dave Grosvenor
Media Technologies Laboratory
HP Laboratories Bristol
HPL-2007-12
February 5, 2007*

photo-album, cast, person recognition, person identification, emphasis, emphasis statistics

This technical report was originally externally published as a defensive publication from an invention disclosure. However it was decided to make the paper easily available to the HP technical community. A demonstration of the work has been produced and the work progressed. (see http://w3.hpl.hp.com/people/dag/cast/cast_overview.htm).

A photo-album is generated using the specification of a "cast" to direct the emphasis of particular people.

It is a simple means of exploiting some additional semantic information obtained using person identification, and can easily be tuned to present many generic stories involving people without any deep semantic knowledge of the actual story.

The establishment of a cast allows people-oriented variations of the photo-album that are understandable and controllable by a user. Furthermore manually establishing the "cast" is a powerful mechanism for controlling the presentation and provides a psychologically important step for a user to establish ownership over the generated presentation.

* Internal Accession Date Only
Approved for External Publication

The use of a cast to generate person-biased photo-albums

Dave Grosvenor
26th July 2006

Introduction

This invention generates a photo-album using the specification of a "cast" to direct the emphasis of particular people or objects through the use of object recognition or identification techniques ([1][2][3][4][5]). The photo-albums are composed from an input set of photographs and videos.

We will use the terms "person" or "people" and "face" to refer to a member or members of the cast and the primary view of an object, but it should be understood that the invention applies to the identification of arbitrary objects that can occur in the cast. This is done because people are the most important objects occurring in peoples photographs

The underlying "story" of most events is about people, or at least a people-oriented slant produces satisfactory results. Most events photographed are about people. This invention can easily be tuned to present many generic stories involving people without any deep semantic knowledge of the actual story.

This invention provides photo-album variations that are understandable and controllable by a user. It gives a simple and intuitive means of controlling the photo-album generated. Manually establishing the "cast" is a powerful mechanism for controlling the presentation and provides a psychologically important step for a user to establish ownership over the generated presentation.

Emphasis

Within a photo-album several techniques are used to emphasize particular pictures or objects.

- The selection of the photographs presented in the album emphasizes the pictures or objects in the collection through presence and distributions of particular objects.
- Electronic photo-albums can be generated in template-styles where pages presented the images at several different sizes.

This allows the emphasis of particular images by varying their size and number on the page.

- The selection of the set of pictures (or objects) to show on a page. The other objects generate competition for the visual attention of a viewer of the album. So both the number of the

pictures per page and relative size of the objects on the page is important. Similar comments apply when pages face each other within the album.
- The spatial position of particular objects within an image and the position of the image on the final page.
- The original images can be cropped around particular objects to generate alternative images in tune with the overall presentation ([7][8]).
- Objects within an image can be emphasized (or de-emphasized) without cropping by blurring the background or other objects ([6]).
- Except for the initial selection of images, these emphasis-techniques are not easily performed when photo-albums were composed only from photographic prints. Thus digital imaging allows the production of better or more dynamic photo-albums.

The cast

The "cast" might be determined automatically from analysis of the input photo-set, or more powerfully it could be determined with some user interaction.

The specification of the "cast" directs the layout algorithm to emphasize particular people (or objects) in the final presentation. The notion of a cast is used to indirectly control the appearance of the photo-album. The cast is used:
- To identify the people to be emphasized.
- To identify relations between people by emphasizing particular groupings of people.
- To emphasize particular spatial configurations of people
- To specify the relative distributions of particular people and groups of people.
- To specify the people and relations emphasized can vary throughout the album. At the beginning of an event we might choose to equally weight each actor to introduce them, before choosing to emphasize the "star" actors. Similarly at the end of an event the entire cast might be shown again. These variations are usually the result of some stylistic parameter for the whole presentation.

These semantic observations of the cast of the album will indirectly control the mix of shots (close-up, medium, long) used for particular people by measuring the emphasis given to particular people in the final album.

A simple measure of the emphasis placed upon a particular person is the sum of the area of their face's whenever they occur in the final photo-album. More complex measures would
- Weight the sharpness and quality of the face image.
- Spatially weight the face according to its position within the photograph itself. Central faces could be weighted more.
- Spatially weight the face according to its position on the final album page. This weight is determined by the template-style which makes certain parts of the page more prominent.
- Take into account the size of the other competing faces on the page.

Analysis

Techniques of object recognition and visual similarity ([1][2][3][4][5]) are used to analyze the input photo-set to measure the emphasis on different people and groupings of people. This analysis gathers statistics of the people or objects identified in the original photo-set recording:
- The combinations of people that occur together.
- The emphasis placed upon particular people by their size and position in the original photo
- The introduction of newly identified people into the presentation.
- The importance of pictures near the start and finish of scene or event boundaries within the original photo-set.

- The unusual pictures in the photo-set.
- The sets of visually similar pictures.

This analysis can use metadata from the larger photo-collection containing it. This makes it easier to both recognise people already known in the collection, and identify useful relationships between these people (such as husband, wife, parent, etc…) that could affect the cast.

This analysis provides a context for the photo-set. The cast assigns roles for people occurring in this context for the photo-set. But the statistics for the object emphasis produced by the cast need not reflect the original context. The emphasis statistics provided by the original photo-context can be modified by either stylistic or manual controls. In the extreme, the cast can produce emphasis on particular actors that is independent of the emphasis present in the input photo-set.
The fidelity to the original context is a creative control that allows different variants to be produced.

- "Who was present" - here we show each person in the cast with equal emphasis despite the input statistics.
- "We were there" - here we show the main actors, but there is also a need to identify the location or context.
- "A child's birthday party" -- Here there is a single main actor and she is likely to occur most frequently in the pictures, although her best friends might be given supporting roles.
- "Watching the school soccer team play" – here there is a set of people in the school team, but also there is another team of distracting objects that could be de-emphasized. Individual parent might want to emphasize their own child.

Emphasis driven layout

Once the cast has been determined we have specified the desired emphasis to be placed both upon particular people, and combinations of people in the photo-album. Now the layout algorithm has to generate an album with an acceptable fit to the desired emphasis. This requires some form of optimisation-like search through the space of potential layouts and the use of the emphasis measure to compare the desired emphasis (given by the cast) with that created by a particular layout.

The problem with graphical design or typographic techniques is that seemingly small differences produce variations that make visually important differences. This creates a combinatorial explosion in the number of possible ways of laying out the album. But equally the optimisation-search would need to be able to discriminate the different (but subtle) variations.

A practical implementation of the layout algorithm would use various assumptions to reduce the search space (such as limiting the window-size of photos on which layout is performed).

In particular, we assume that a template-style that has been determined prior to layout. The template-style provides a set of potential templates that will layout the photo-album with the selected aesthetics. For each page a template defines a number of slots in which arbitrary images can be placed at a particular size. This fixes many of the graphical design decisions, and leaves our layout algorithm with the job of selecting images to place in the slots.

The template-style is responsible for background selection, the use of stock-graphic objects, the use of white-space, the imposition of a grid-structure on the album, the space between images (guttering), the style of titles and captions, etc… The template-style has to introduce variation between different pages of the photo-album whilst binding them together in a complementary fashion.

This invention is particularly suited to using template-styles that vary the size of the photos on the album pages because they allow the layout algorithm to emphasize particular people or objects. A significant choice of emphasis is forced on every page when the template style for the page has a single large photo composed with much smaller photos. This creates a winner-take-all form of emphasis, where a chosen object is greatly emphasized in turn.

References

1. "Pattern Classification"(2nd ed.) by Richard O. Duda, Peter E. Hart and David G. Stork, Wiley Interscience, 680 pages ISBN: 0-471-05669-3
2. "Dynamic Vision: From Images to Face Recognition" By Shaogang Gong, Stephen McKenna, Alexandra Psarrou , 364 pages, Imperial College Press, 2000.
3. "*State-o-fthe-Art in Content-Based Image and Video Retrieval",* Veltkamp, R. C., Burkhardt, H., and Kriegel, H.-P., editors (2001. Kluwer Academic publishers.
4. "*Image Databases: Search and Retrieval of Digital Imagery*", Castelli, V. and Bergman, D., editors (2002). John Wiley & Sons, Inc.
5. "Special issue on content-based multimedia indexing and retrieval", Djeraba, C. et al. (2002).. *IEEE Multimedia Magazine*, 9(2):18.60.
6. "De-Emphasis of Distracting Image Regions Using Texture Power Maps". Sara L. Su, Frédo Durand, and Maneesh Agrawala. In Texture 2005: Proceedings of the 4th IEEE International Workshop on Texture Analysis and Synthesis in conjunction with ICCV'05, pp. 119-124, Beijing, China, October 2005.
7. **"**A visual attention model for adapting images on small displays**",** MSR-TR-2002-125, Liqun Chen; Xing Xie; Xin Fan; Wei-Ying Ma; Hong-Jiang Zhang; Heqin Zhou, November 2002.
8. "Automatic thumbnail cropping and its effectiveness ", Suh, B., Ling, H., Bederson, B. B., and Jacobs, D. W., *ACM Conference on User Interface and Software Technolgy (UIST 2003)* (2003), 95–104.