Using the emphasis statistics of a cast for the automatic layout of a photo-album

Dave Grosvenor, Susanne Klein
Media Technologies Laboratory
HP Laboratories Bristol
HPL-2007-52
April 18, 2007*

This paper describes a simple and intuitive method for controlling the automatic layout of a photo-album. It empowers a consumer-user by taking advantage of their knowledge of the importance of particular people (the cast) in their story. Currently automatically detected faces are manually tagged. A cast of identified people is established which specifies the visual impact (emphasis) of a person desired in the presentation. This provides a powerful means of controlling the presentation without any direct manipulation. It enables a user to generate different person-biased variants from the same photo-collection and so tailor a presentation to a particular audience.

 The idea is to allow the user to control the overall intended emphasis given to each member of the cast. This target emphasis has to be achieved within the aesthetic constraints established by a user selected style. For each potential photo-selection, the style is responsible for enumerating the possible template layouts. This enumeration reduces the layout-problem to one of: selecting from a set of templates, and deciding which photo to put in a slot of the template. This allows the use of a search to identify the best combination of template and photo-selection. The search discriminates candidate layouts by comparing the emphasis measured in a particular layout against the desired target emphasis.

Using the emphasis statistics of a cast for the automatic layout of a photo-album

Dave Grosvenor
Susanne Klein
Hewlett-Packard Labs
20th March 2007

Abstract

This paper describes a simple and intuitive method for controlling the automatic layout of a photo-album. It empowers a consumer-user by taking advantage of their knowledge of the importance of particular people (the cast) in their story. Currently automatically detected faces are manually tagged. A cast of identified people is established which specifies the visual impact (emphasis) of a person desired in the presentation. This provides a powerful means of controlling the presentation without any direct manipulation. It enables a user to generate different person-biased variants from the same photo-collection and so tailor a presentation to a particular audience.

The idea is to allow the user to control the overall intended emphasis given to each member of the cast. This target emphasis has to be achieved within the aesthetic constraints established by a user selected style. For each potential photo-selection, the style is responsible for enumerating the possible template layouts. This enumeration reduces the layout-problem to one of: selecting from a set of templates, and deciding which photo to put in a slot of the template. This allows the use of a search to identify the best combination of template and photo-selection. The search discriminates candidate layouts by comparing the emphasis measured in a particular layout against the desired target emphasis.

Introduction

The automatic creation of photo-albums is temptingly appropriate, for a consumer-user who is both time poor, and has little understanding of photo-album aesthetics. The aim has been to produce something that looks good with the photos the user selected. The risk is that the album is regarded as something the computer generated rather than something they created. The user needs to feel the result is individual and special. This work empowers the user by taking advantage of their knowledge of the importance of particular people (the cast) in their story. A cast of identified people is established which specifies the visual impact (emphasis) of a person desired in the presentation. It enables different variants of the same photo-collection to be created easily, allowing a presentation to be tailored for a particular audience.

The idea is to allow the user to control the overall target emphasis given to each member of the cast. A target emphasis for each person in the cast is obtained by weighting their emphasis measured in the input photos. This requires a measure for the emphasis given to the cast by a set of photos.

When a cast control is not used, a photo-album can emphasize particular people by:
- The selection of the particular photos presented, which contain both different people and different shot types of people (close-up, medium, long).
- The presentation of the selected photos at different sizes within the album.

Such selection and preferential presentation of photos are important creative decisions that are taken for each photo, whilst considering both the overall intended effect and stylistic constraints. Expressing such a preference for each photo is a difficult and time consuming task, which gets worse as the number of photos increases. Editing the appearance using direct manipulation of the album itself to meet these goals is more difficult. Establishing the cast allows both selection and preferential presentation of photos (style permitting) to be controlled without the need to inspect each photo.

This is a simple but powerful approach which can tailor a presentation, to a particular audience to obtain different variations of person-biased presentations:

- Representative selection – where the target emphasis statistics of each member of the cast is determined by the emphasis given to them in the input photo-collection. The emphasis given to particular people in the collection could be reproduced in the album whilst presenting a subset of the photos.
- Manipulating the relative impact of different people – this can be used to force the relative proportions of different people, perhaps to redress imbalances introduced because one family member takes most of the photos.
- More simply an album stressing a particular person (or suppressing someone) could be generated.
- Different variations can be produced for different target audiences from the same photo-collection.
    - i.   A teenage daughter would use the controls differently for showing her friends the family holiday pictures than her middle-aged father.
    - ii. For multi-family re-unions or gatherings a different version could be created for each family.
    - iii. Photos taken of a school soccer game could have a different version for each parent.

Our cast-based approach [1][2] to automatic layout of photo-album differs from existing work because:

- It provides the user with a method of controlling the visual impact of a person in a presentation without directly manipulating the appearance.
- It uses analysis of the image content to perform the layout (unlike [3]).
- Some approaches [3] allow a user to control the layout of photos at different sizes, by giving a target relative area for each photo. Our cast-based approach controls the relative sizes of photos indirectly through setting the visual importance of each person without the need to inspect every photo. Photos are only presented at different sizes if this is permitted by the current user-selected style.
- Identified faces are used to control the visual impact given to a person in a presentation, rather than for determining event boundaries and pagination using either temporal or face-based clustering [4][5].
- It intimately combines presentation and selection. A photo is selected because it has a part to play in the presentation. Other approaches either present every input photo [3][4][5], or separate photo-retrieval from the presentation. For example, they might retrieve all the pictures of a particular person before presenting those selected [6].

Establishing the cast

Manually establishing the cast is a simple and important creative control that empowers the user. This interface is used to create target emphasis statistics to specify the overall presentation. The target emphasis statistics are used to indirectly control the mix of shots (close-up, medium, long) in the generated presentation. The envisaged interface, in Figure 1, is something like an audio-mixing desk controlling the emphasis of different people. Cast-based presentation is not just about identifying the people. It is deciding who is important to the story. This is a creative decision. Currently automatically detected faces are manually tagged with a person's identity.
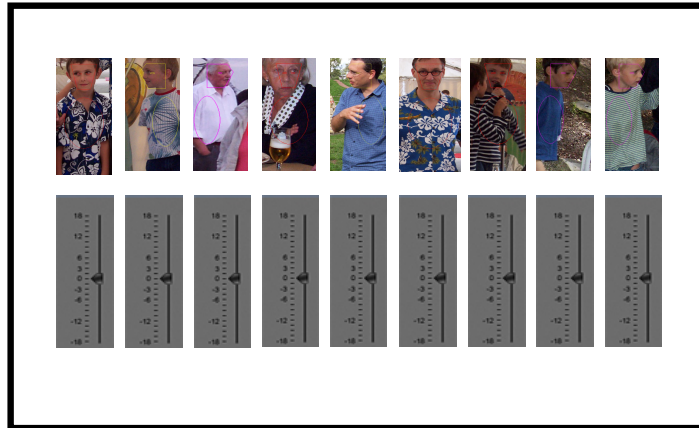

Figure 1 – An envisaged interface to establish the importance of members of the cast.

Providing a control to our target user is more difficult than simply using a constraint as a heuristic to generate a better presentation. A control needs to be responsive and useful over a whole range of values. An important prerequisite for such a presentational control is:
- o A target or specification and;
- o A means of measuring the extent to which a given presentation satisfies the specification.

The current cast-control mechanism needs improvement, as it is possible to request an emphasis target that cannot be fulfilled. The emphasis targets that can be achieved are the limited by the contents of the photos and the style chosen to present the photos. The cast control works best when either the template-style allows photos to be presented at a range of sizes, or the there is a mix of shots of different people available for selection.

Measures of emphasis

In our approach to cast-based layout:
- The emphasis in a photo-collection is used to set both the overall target emphasis statistics selected by the user, and indirectly the target emphasis statistics for each page.
- A search generates potential page layouts (applying a photo-selection to the slots of a template). The emphasis given by a potential page-layout is then compared with the target emphasis to choose a particular page-layout.

This requires a measure for the visual impact (emphasis) given to the cast of identified people in a photo, a collection of photos, an album page layout, and the whole album. The simple measure used for each of them is the relative area of the persons face compared to the display area.

Leon   Nicolai

Figure 2 The emphasis of two detected people (Leon and Nicola) is the area of their faces divided by the area of the photo.

For a photo, this is the area of the face divided by the image area. Currently the bounding-box returned by the HP face-detector is used to measure of the face area (see Figure 2, but more precise calculations should be used). A close-up uses about 5% of the image area to produce an emphasis of 0.05 for the person. Different shot types would have different emphasis for the person. Figure 3 shows the variation in this emphasis measure of a person (Leon) over the whole photo-collection.

This is a very simple measure of the visual impact of a person. The position within the photo, the sharpness/blur of the persons face in the photo, the presence of other people, the saliency of the persons face in the photo, their sex-appeal, make-up, the highlighting of faces, could all be taken in to account or manipulated.
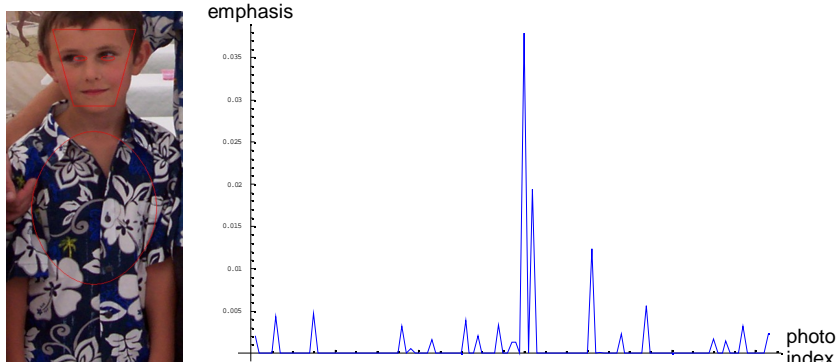
Leon

Figure 3 The emphasis per photo of a particular person (Leon) over a photo-collection of over 120 photos.

For a photo-collection, the average emphasis given to a photo is used (sum the emphasis from each photo, then divide by the number of photos). This measure allows a single face of a person to produce the same emphasis as four images of the same person that are half this height. The search for a layout is allowed to achieve its target emphasis by either way. More sophisticated emphasis statistics can be used to control the mix of shot types used to achieve a target.

Figure 4 shows a photo-collection from a holiday visiting remote family, and shows people from several related families (nearly 40 people in this cast). The emphasis of each person in the collection is small (compared to the emphasis produced when someone occurs in a photo) because they do not occur in every photo.
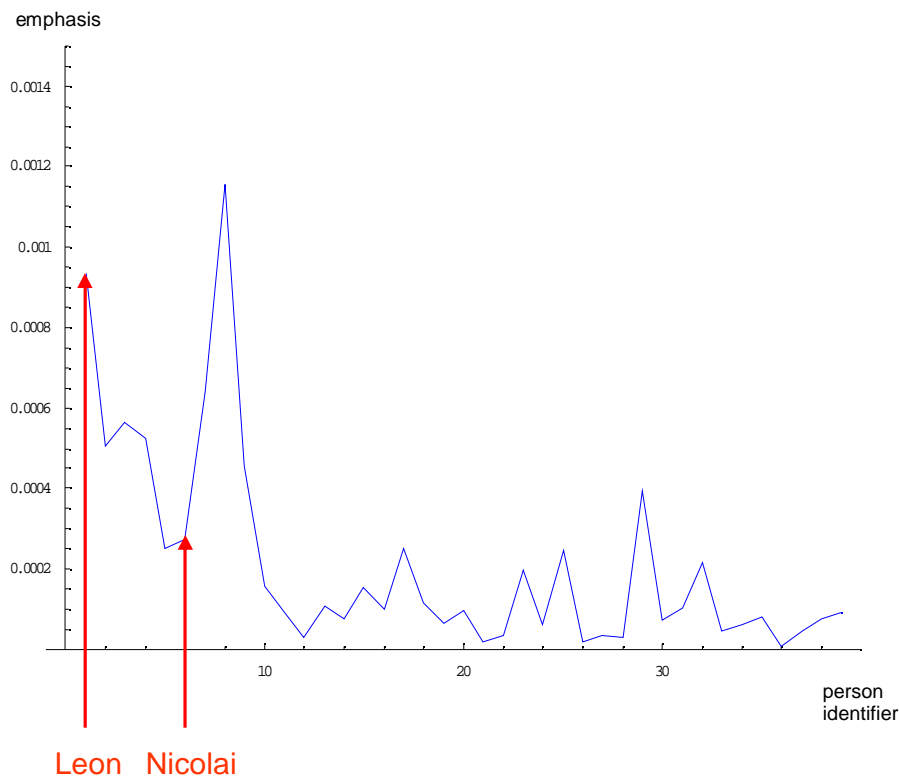

Figure 4 The emphasis of a cast (nearly 40 people) measured in a photo-collection

The emphasis produced by an album page (or facing pages) is obtained by summing the total area from all the faces of a person on the page and dividing it by the area of the page. This treats at album page in the same way as a single photo. A certain bias in the measure reducing the emphasis on a page is introduced by the use of white-space or non-photo-space. More sophisticated emphasis measures take into account spatial position, and the use of graphics to enhance the visual impact of a photo or person on the page.

Figure 5 shows an album page layout and the emphasis it produces for the cast.

The emphasis produced by a whole album is the average emphasis given to the album pages. The emphasis produced for each page is summed and divided by the number of pages. This is the same approach as treating an album like a sequence of photos. More sophisticated approaches take into account the greater prominence of initial and final parts of the album.
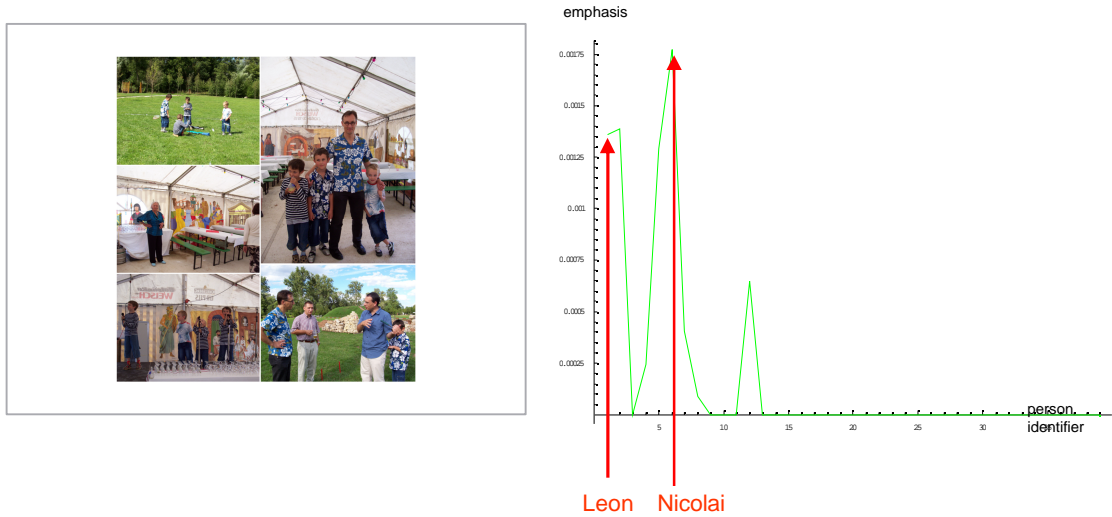

Figure 5 The emphasis of the cast produced by an album page

In Figure 6, the target emphasis for the whole album of each person (red) is derived by weighting the emphasis in the input photo-collection (blue when different from the red). The emphasis produced by the album layout generated is shown in green. For some members of the cast the album generated exceeds the target, but on others the target is not met. Any album generated has to trade performance between the different cast members.
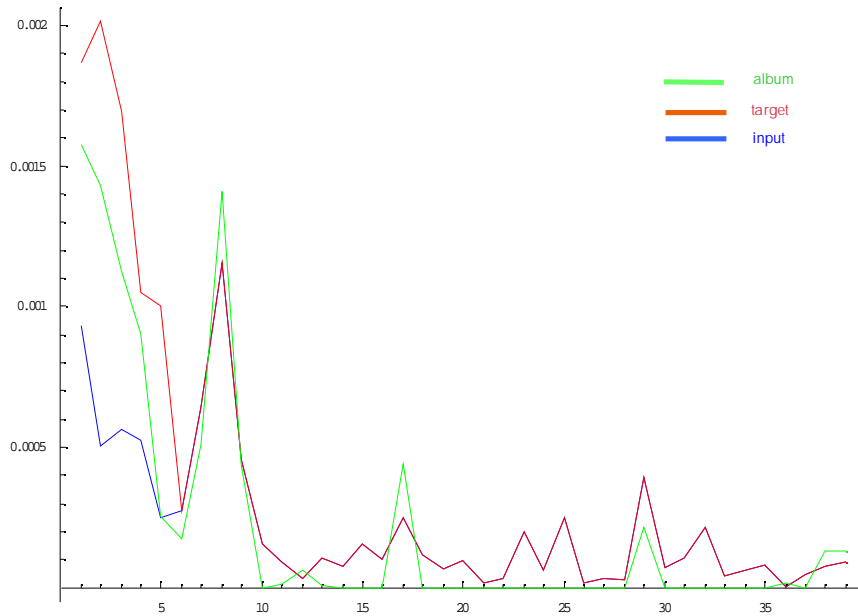

Figure 6 Comparing the emphasis produced by an album (green) against its target (red)

Emphasis targets for a page

The target emphasis statistics for the overall album are be decomposed into target emphasis statistics for each page (or facing pages) using some simple heuristics:

- A set of candidate photos for the page is determined using some simple heuristics (taking into account the target number of pages, the number of photos per page permitted by the set of style-templates, the desired ratio of presented photos to input photos).
- The same user determined weights obtained by establishing the cast are applied to the emphasis in the set of candidate photos.

These heuristics for obtaining emphasis targets per page reduce the complexity of the combined search over possible templates and photo-selections.

The target emphasis derived for each page is dependent on the emphasis available in the set of candidate photos. This makes it more likely that the emphasis targets are achievable. However if the emphasis targets set for a page are not achieved then the page targets of following pages are adjusted by modifying the weights.

Photo-album styles

This section examines the responsibilities of the designer of a photo-album style to be used with our cast-based automatic layout. Our approach works within the aesthetic constraints imposed by a style, and discriminates different layouts permitted by the style using the importance of members of the cast.

Families of complementary templates have been used to implement many different styles, but the design principles they encode are not explicit. Each design is visually complex (controlling the white-space, spatial-balance, grid alignment of the text, title and graphics). Developing automatic layout algorithms embodying such aesthetic design rules for such a number of styles is difficult. The use of templates avoids the problem of characterising a design from first principles, but it introduces other problems.

Our styles characterise a design through the set of templates that can be used to layout an album page using a particular sequence of photos. This set of templates is dependent both on the photos, and on the relation of this particular album page to the whole.

A designer has to enumerate the templates that can be used for a particular album page with particular classes of photos.

- Different styles of left and right pages, complementary facing pages, different styles for initial and final page.
- Enforcing the desired variation of page layouts in the album by making choice dependent on the templates selected for other pages.
- Simple templates take a sequence of photos with particular aspect ratios, compositing the photos with the background and graphics to produce a page layout. They often require the photos to have particular aspect ratios because otherwise they interfere with the design. Thus the set of template is usually dependent on the aspect ratios of the photos.
- If the style is to cope with a range of aspect ratios then the design has to anticipate them. Unfortunately this creates a combinatorial explosion in the number of templates that have to be enumerated.

- The only templates currently supported are built using hierarchical page structures [3] that can specify template-layouts with some independence of the aspect-ratios of the photos. Hierarchical page structures layout photos with arbitrary aspect ratios, but the result is not always desirable.
- A smaller range of aspect ratios and the use of cropping is a pragmatic approach.
- Similarly if the photos are to be presented at different sizes on a page then this also has to be anticipated by the designer.
- Applying the wrong photos to a template can produce poor results. Currently the layouts are not dependent on the image content of the photos, but some dependence is useful to require:
  - A minimum size for faces in certain slots.
  - Selected photos to have complementary colours to the graphics used with the template.
  - Certain types of photos to be placed in certain slots.
- A cartoon ordering is imposed on the photos that can be placed in a template. This convention imposes a temporal ordering on the photos on the page with time increase top to bottom, and left to right. More generally to allow some slots not to be ordered against other slots, a partial ordering is defined between slots of a template. This reduces the number of possible selections that need to be explored by a search.

Searching for a page layout

For each potential photo-selection, the style is responsible for enumerating the possible template layouts. This enumeration reduces the layout-problem to one of: selecting from a set of templates, and deciding which photo to put in a slot of the template. This allows the use of a search to identify the best combination of template and photo-selection.

Currently a brute-force over all possible templates and photo-selections is performed. But faster searches getting a good rather than the best pairing (using a genetic algorithm) could be used.

The search discriminates candidate layouts by comparing the emphasis measured in a particular layout (a) against the desired target emphasis (b).

$$d(a,b) = \sum_{p \in People} w_p \big(emphasis(a, p) - emphasis(b, p)\big)^2$$

Different weights can be used to control the trade-off in the emphasis between different members of the cast.

Person-biased variants

Two examples of different person-biased variants created from the same set of candidate photos are discussed.
- Figure 6 shows the set of candidate photos used to create both person-biased variants.
- Figure 7 shows a variant where the target emphasis is representative of the overall photo-collection. It shows both the generated album page from the set of candidate photos, and three superimposed graphs:- showing the target emphasis (red) used compared with the emphasis produced by the generated album page (green). The blue graph shows the emphasis of the set of candidate photos
- Figure 8, shows a variant where illustrates the target emphasis (red) that stresses one member of the cast. It shows both the generated album page, and three superimposed

graphs: showing the target emphasis (red) used compared with the emphasis produced by the generated album (green). The blue graph shows the emphasis of the candidate set of photos.

Figure 6 The set of candidate photos used for generating both person-biased variants of an album page.
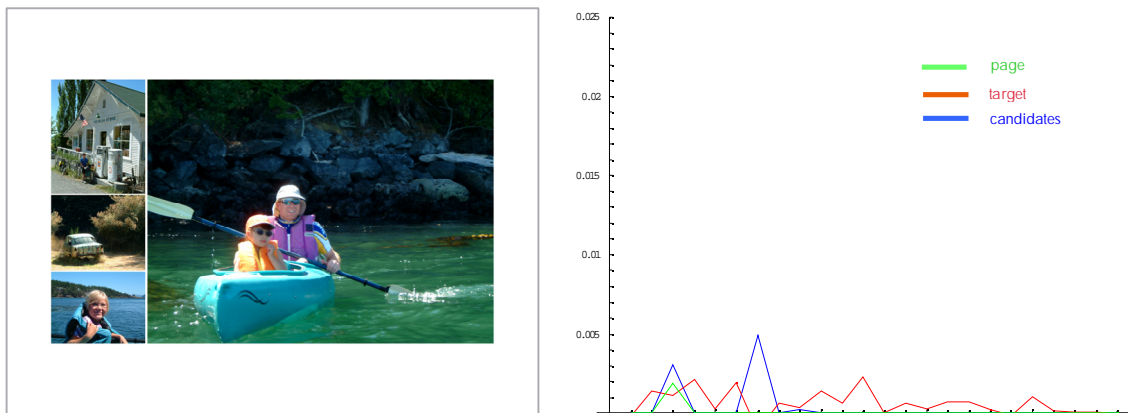

Figure 7 The album page generated by a *representative selection* with a graph showing the target emphasis (red) compared with the emphasis produced by the album page (green)
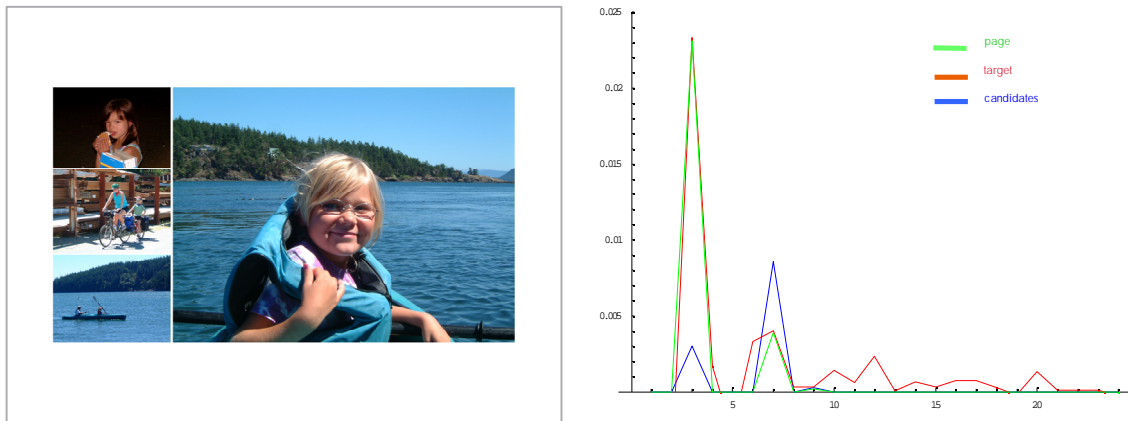

Figure 8 The album page generated by *stressing one person* with a graph showing the target emphasis (red) compared with the emphasis produced by the album page (green)

Conclusions

1) An intuitive but powerful user interface for controlling the automatic layout of photo album has been described.
2) The control is implemented using emphasis statistics to characterise the desired album and to discriminate the different album layouts generated during a search.
3) Different person-biased variants produced using the control have be illustrated.
4) Further work is needed to improve the responsiveness of the control and a better search procedure is needed.
5) This control can be applied to many other forms of presentations than photo-albums and was initially conceived as a control for photo-video slideshows [1].

References

[1] "The use of a cast to generate person-biased video presentations", Dave Grosvenor, Susanne Klein, 26th July 2006, HP defensive publication. HP Labs Technical Report HPL-2007-11
http://www.hpl.hp.com/techreports/2007/HPL-2007-11.html

[2] "The use of a cast to generate person-biased photo-albums", Dave Grosvenor, 26th July 2006, HP defensive publication. HP Labs Technical Report, HPL-2007-12
http://www.hpl.hp.com/techreports/2007/HPL-2007-12.html

[3] "Adaptive photo-collection page layout", C. Brian Atkins,
2004 International Conference on Image Processing,
Oct. 2004, pp 2897- 2900 Vol. 5, ISBN: 0-7803-8554-3
http://www.hpl.hp.com/research/isl/layout/icip04_abstract.pdf

[4] "Automatic face-based image grouping for albuming", Madirakshi Das, Alexander C. Loui, IEEE international conference on systems man and cybernetics, 2003

[5] "Automated event clustering and quality screening of consumer pictures for digital albuming", IEEE transactions on multimedia, v0l. 5, n.o. 3, september 2003.

[6] "Motivating Annotation for Personal Digital Photo Libraries: Lowering Barriers While Raising Incentives", Univ. of Maryland Technical Report HCIL-2004-18, January 2000