Smart Photo Sticking

Sebastiano Battiato[1], Gianluigi Ciocca[2], Francesca Gasparini[2], Giovanni Puglisi[1], and Raimondo Schettini[2]

[1] Dipartimento di Matematica ed Informatica
Università di Catania
viale A. Doria 6, 95125 Catania, Italy
[2] DISCo, Dipartimento di Informatica Sistemistica e Comunicazione
Università degli Studi di Milano-Bicocca,
via Bicocca degli Arcimboldi 8, 20126 Milano, Italy
{battiato, puglisi}@dmi.unict.it,
{ciocca, gasparini, schettini}@disco.unimib.it

Abstract. Smart photo sticking is a novel strategy to automatically arrange a collection of photos in a pleasant collage. The proposed approach improves previous solutions both considering a self-adaptive image cropping algorithm, exploiting visual and semantic information, and introducing an optimization process based on a genetic algorithm. Preliminary results confirm the effectiveness of the proposed strategy on a heterogeneous collection of non professional photos.

Keywords: Image summarization, saliency region, genetic algorithm.

1 Introduction

Images are playing a more and more important role in sharing, expressing and exchanging information in our daily lives. Now we all can easily capture and share personal photos anywhere and anytime.

Generating a collage that summarizes a group of pictures could be trivial if realized manually 1, but developing an automatic tool is intrinsically difficult. A simple technique for image arrangement is page layout 2 that tries to cover the canvas area with no overlap without considering (or distinguishing) the relevant regions of each input image. On the other hand most previous image summarization works are mainly based on content based techniques (3, 4) to provide a high-level description of a set of images. The CeWe colourbook album software 5, does a lot of cropping, salience recognition and collaging; it detects snapshots which are out of focus and also those which are over-exposed, under-exposed or double, though on photo album and not on poster. A different approach for image summarization is presented in (9, 10) where the collage is assembled joining subsets of each picture together by using ad-hoc techniques (i.e. Poisson Editing, etc.) to hide the joins between input images. An other interesting way of viewing photos on a computer is Microsoft Photosynth 11; it takes a large collection of partially overlapping photos of a place or an object, analyzes them for similarities, and then displays them in a three-dimensional space. Some further approaches (6,7,8) have been recently proposed by obtaining impres-

N. Boujemaa, M. Detyniecki, and A. Nürnberger (Eds.): AMR 2007, LNCS 4918, pp. 211–223, 2008.

sive results just considering ad-hoc heuristics for both image analysis and layout positioning.

One of the most successful attempt to manage this kind of situation is Picture Collage described in 12, where the authors propose a system to arrange groups of pictures inside a canvas with possible overlay, minimizing the occlusion of salient regions of each involved image. In 12 the image arrangement is formulated as a Maximum a Posterior (MAP) problem such that the output picture collage shows as many visible salient regions (without being overlaid by others) from all images as possible. Moreover, a classic Markov chain Monte Carlo (MCMC) method is designed for the optimization.

We propose two different improvements with respect to the work presented in 12. First of all, the detection of the saliency region has been performed by applying a novel self-adaptive image cropping algorithm which exploits both semantic and visual information. Semantic information relates to the automatically assigned image categories (landscape, close-ups, …) and to the detection of face and skin regions, while visual information is obtained by a visual attention model 13 that has been developed in order to find a salient image region to be cropped, and visualized on small displays. In this work, the cropping area is used to drive the photo sticking.

The second improvement is related to the different optimization criterion used. We have implemented a genetic algorithm able to capture the different constraints derived directly from the typical good layout that a collage of photos must satisfy. Preliminary results confirm that the fitness we have designed is able to reach a good solution in almost all cases.

The paper is organized as follows. Next Section briefly summarizes the self-adaptive cropping system used to locate the saliency region inside each image. Section 3 is devoted to describing the main underlying ideas of the proposed genetic algorithm. Preliminary results are presented in the next Section while Section 5 closes the paper tracking directions for future works and research.

2 Self-Adaptive Image Cropping

As stated before, we use a self-adaptive image cropping algorithm to detect the relevant region within an image. This information is then fed to the algorithm responsible for the photo arrangement. Most of the approaches for adapting images only focused on compressing the whole image in order to reduce the data transmitted. Few other methods use an auto-cropping technique to reduce the size of the image transmitted 14, 15. These methods decompose the image into a set of spatial information elements (saliency regions) which are then displayed serially to help users' browse or search through the whole image. These methods are heavily based on a visual attention model technique that is used to identify the saliency regions to be cropped.

Figure 1 shows the flow diagram of the algorithm we have developed. The images are first classified into three broad classes, that is, "landscape", "close-up", and "other". The classification is based on the use of ensembles of decision trees, called decision forests. The trees of the forests are constructed according to CART (Classification and Regression Trees) methodology 16. The features used in this classification process are related to color, texture, edge and composition of the image 17, 18. Then, an ad-hoc cropping strategy is applied for each image class.

The detection of the relevant region depends on the image content. The three broad classes have been chosen so that they cover the main groups of images that we may find in any collection of photos. The classification of the images allows us to build a detection strategy specific for each image class. The effectiveness of the strategy is maximized by taking into account the properties of the image and focusing our attention to some objects in the image instead of others. A landscape image, for example, due to its lack of specific focus elements, is not processed at all: the image is regarded as being wholly relevant. A close up image, generally, shows only a single object or subject in the foreground, and thus, the relevant region should take into account only this, discarding any region that can be considered as background. In the case of an image belonging to the other class, we are concerned if whether it contains people or not: the cropping strategy should prioritize the selection of regions containing people.

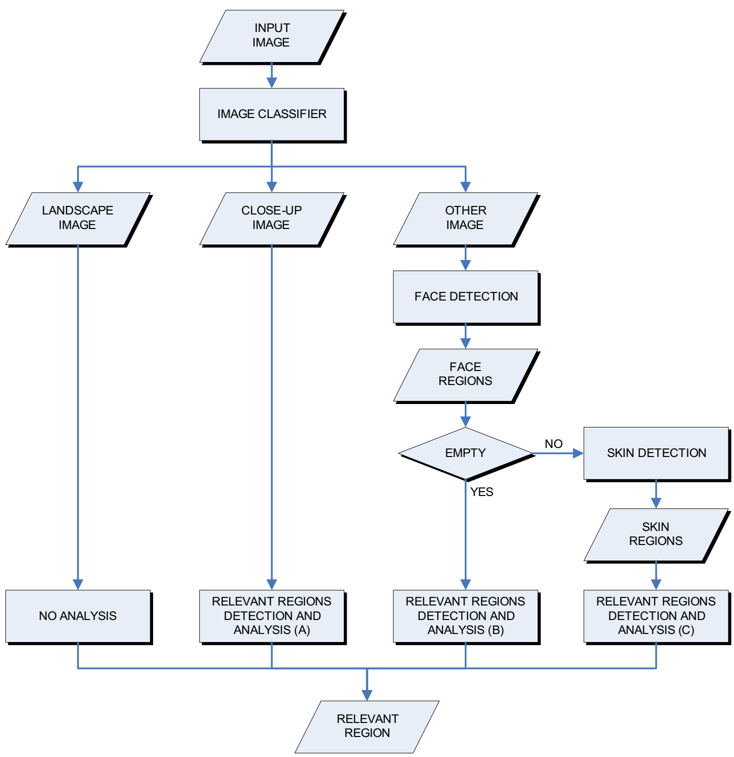

Fig. 1. The flow diagram of the proposed algorithm

Landscape images. In the case of landscape images, no cropping is performed. We adopt this strategy because landscape images usually do not present a specific subject to be focalized. Examples of landscape images are reported in Fig. 2.

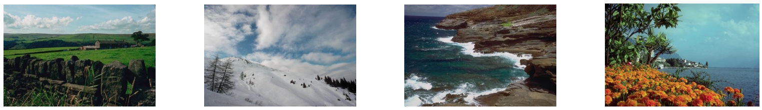

Fig. 2. Examples of “Landscape” images

Close-up images. For close-up images here we define a new procedure which we called “Relevant regions detection and analysis (A)” in Fig. 1:

a. A saliency map is generated based on the Itti-Koch visual attention model 19. Visual attention facilitates the processing of the portion of the input associated with the relevant information, suppressing the remaining information.
b. The saliency map is automatically binarized in order to identify saliency regions. The regions with areas smaller than a threshold which are a function of the area of the larger region are discarded.
c. A single relevant region is obtained, considering the bounding box that includes all the saliency regions previously identified.
d. The image is then cropped with respect to this region.

Other images. A face detector inspired by the Viola and Jones one 20 is applied to distinguish between images with and without faces. The detector is composed of a chain of weak classifiers trained by the Ada-boost algorithm.

For images without faces we designed the “relevant region detection and analysis (B)” strategy:

a. Same as point a. of the close-up case.
b. Same as point b. of the close-up case.
c. The most salient region identified in point b. is now considered as the relevant region.
d. The image is then cropped with respect to this region.

For images with faces, we designed the “relevant region detection and analysis (C)” strategy:

a. Same as point a. of the close-up case.
b. A skin color map is also computed. For the skin detector we adopted an explicit skin cluster method based on the YCbCr color space, where the boundaries of the color skin cluster were chosen to offer high recall in pixel classification 21.
c. The saliency map, the skin color map and the face regions are then combined together to form a global map, used to locate the most relevant region.
d. The image is then cropped with respect to this region.

For both *close-up* and *other* images, the borders of the final cropped region are enlarged to include the relevant area better.

In Figures 3-5 examples of the final crop region for these different classes of images are reported.

Fig. 3. Relevant regions selected within some of the "close-up" images

Fig. 4. Relevant regions selected within some of the "Other" images. No faces are present or detected.

Fig. 5. Relevant regions selected within some of the "Other" images containing faces

3 Photo Arrangement

In order to obtain a good photo sticky the following properties should be considered 12:

- salience maximization, to show in the canvas as many important region as possible;
- blank space minimization, to use all the canvas without holes;
- salience ratio balance, to reach a stable ratio balance (percentage of visible salient region).

To satisfy these properties the problem can be formulated as an optimization problem but, due to its complexity (NP complete), only an advanced (smart) solution has to be designed. For this reason we have chosen a genetic algorithm: an optimization and search technique based on the principle of genetics and natural selection. An initial population, usually randomly selected, of possible solutions evolves toward a better solution. In each step some population elements are stochastically selected based on their fitness (the function to be optimized), and new elements are created through some techniques inspired by evolutionary biology (mutation, crossover). Genetic algorithms have found application in many fields 22: computer science, engineering, economics, chemistry, physics, etc.

Given N input images I_i $i=1, \ldots, N$ and the corresponding saliency maps s_i $i=1, \ldots, N$ the final goal is devoted to arrange such pictures in a canvas C. The canvas is rectangular (with the classic aspect ratio set to 4/3) and its size is set to that its area is about half of the total area of all input images. Each image I_i, in the picture collage, can be labelled as a triplets $\{c_i, o_i, l_i\}$ where c_i is the 2D spatial coordinate of the center, o_i is the orientation angle and l_i is the placement order of the image in canvas.

In order to properly encode salience maximization, blank space minimization and salience ratio balance, we have modelled the fitness function (to be minimized) as follows:

$$Fitness(\overline{A_{occ}}, \overline{B}, V) = e^{(\overline{A_{occ}} + \overline{B} + V)} \ . \tag{1}$$

where:

- $\overline{A_{occ}}$ is the normalized sum of occluded saliency regions defined as:

$$\overline{A_{occ}} = \frac{A_{occ}}{A_{\max}} \ . \tag{2}$$

$$A_{occ} = A_{\max} - A_{vis} \ . \tag{3}$$

$$A_{vis} = \sum_i s_i^{vis} \text{ where } s_i^{vis} \text{ is the visible part of the saliency region } s_i \ . \tag{4}$$

$$A_{\max} = \sum_i s_i \ . \tag{5}$$

- $\overline{B}$ is the normalized sum of canvas uncovered regions defined as:

$$\overline{B} = \frac{Area(B)}{Area(R_c)} \ . \tag{6}$$

$$B = R_c - \bigcup_i R_i \ . \tag{7}$$

where R_i is the bounding rectangle of picture I_i and R_c is the canvas bounding rectangle

- V is the variance of saliency ratios:

$$V = \frac{1}{N}\sum_i (r_i - \bar{r})^2 \text{ where } r_i = \frac{s_i^{vis}}{s_i} \text{ and } \bar{r} \text{ is the } r_i \text{ mean value .} \tag{8}$$

Standard crossover and mutation operators cannot be used directly because each layer order is a permutation: an unique layer index must be assigned to each image. To simplify the problem we initially fix, a layer order $\{l_1, l_2, \ldots, l_n\}$ $l_i \neq l_j \quad i \neq j$, based on the following consideration:

- Let $ns_i = Area(I_i)- s_i$ be the not saliency region of image I_i. Some saliency region occlusions can be avoided positioning the pictures with high ns_i values just below images with low ns values. The ns_i regions can occlude saliency regions and hence they slow function optimization.

The initial layer order is assigned by sorting the input images according to the ns_i values in descending order. In order to speed-up the overall process, the initial population is defined as follows:

- divide the canvas into N* rectangles (with N*>N);
- select N centers c_i* of rectangular blocks;
- sample c_i from a normal distribution with mean c_i* and variance 1/6 of rectangular width and height respectively;
- sample o_i (image orientation) from a normal distribution with zero mean and variance $\pi/27$.

Genetic optimization is realized by using standard approaches in the field. In particular we have used default crossover and mutation algorithms provided by Genetic Toolbox functions of MATLAB 7.

4 Experimental Results

In order to test our solution we have used some image databases obtained by typical consumer usage. In particular, we present preliminary results obtained by considering two different experiments involving 9 and 10 images respectively. The image resolution is about 5MPixel. In both experiments we set the population size to 20 and the generation number to 200.

Figure 6 reports the values of the fitness function with respect to the number of involved generation.

Figures 7 to 10 show the final output by showing also the overall results in terms of saliency region occlusion with respect to the considered layer ordering, canvas usage and image cropping. The final results are promising, in both cases the sum of visible saliency region is about 90% and the overall covered canvas regions are around 99%

of the total. Table 1 reports the final values of $\overline{A_{occ}}$, $\overline{B}$ and V after the optimization process. The timing performances of the system can generate picture collage in about 3 minutes; in this phase we are mainly interested in the effective convergence of the iterative process.

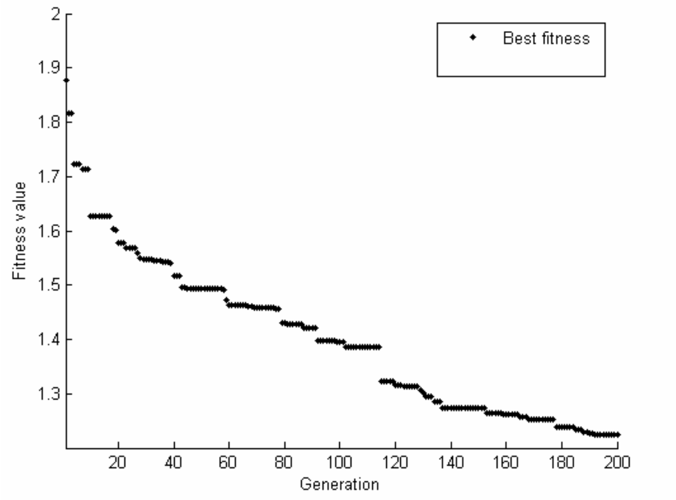

Fig. 6. *Test1* best fitness function

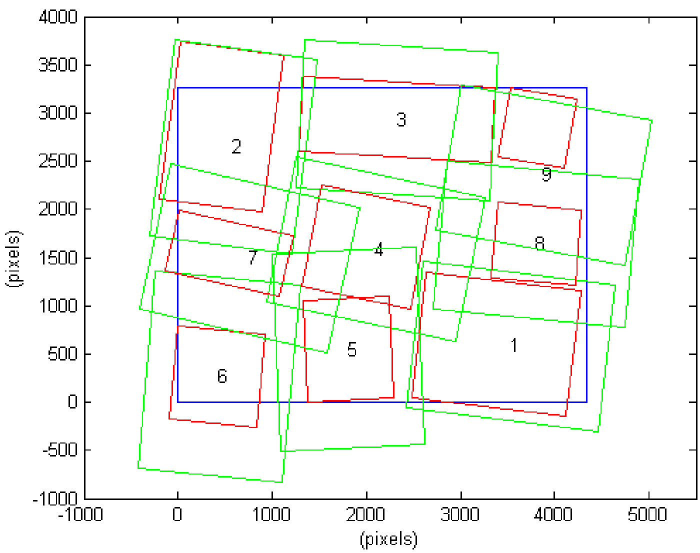

Fig. 7. *Test1* final output in terms of saliency regions *(red)*, image borders *(green)*, layer order and canvas *(blue)*

Fig. 8. *Test1* picture collage

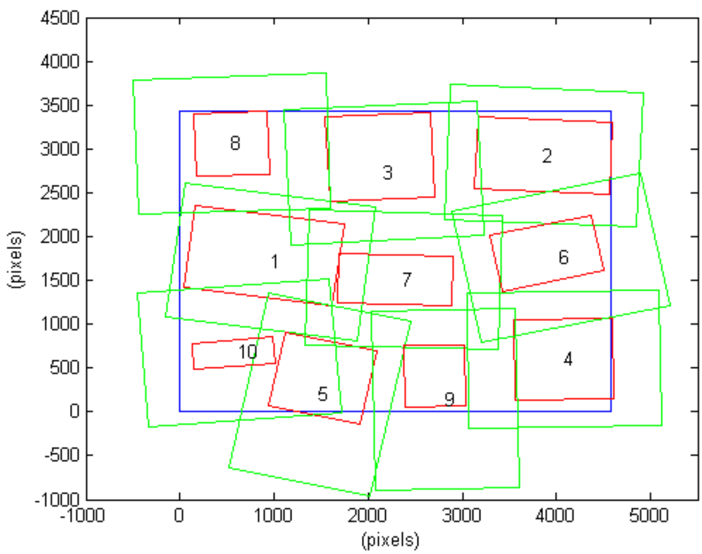

Fig. 9. *Test2* final output in terms of saliency regions *(red)*, image borders (green), layer order and canvas *(blue)*

Our approach works very well for collages of medium size (10-12 photos) finding good solutions quickly. However it lack of scalability in terms of execution time; especially for large collages, the number of needed generation could be very high.

Fig. 10. *Test2* picture collage

Figure 11 shows a comparison between a photo collage which uses the saliency regions detected by the auto cropping algorithm and a photo collage which does not. As can be seen, the disposition of the photos where the saliency regions are used, is more appealing since all the photos are clearly visible and the relevant information is retained in the collage. On the contrary, without the saliency regions, several photos are totally or partially hidden by others and subjects are cut out from the collage. Figure 12 shows a collage which includes images from the landscape class. Since images belonging to this class are considered wholly relevant, they are shown in the foreground forcing the disposition of the other images around them. In order to create a more appealing collage, we plan to introduce rules to cope with landscape images.

Fig. 11. Picture collage comparison, with (left) and without (right) saliency regions

Figure 13 shows a first result where we have forced landscape images to appear in the background instead of the foreground. The improvement is clearly visible.

Fig. 12. Picture collage containing images from the "landscape" class

Fig. 13. Picture collage containing images from the "landscape" class. A rule has been introduced to force these images to appear in the background.

Table 1. Final values of $\overline{A_{occ}}$, $\overline{B}$ and V for *test1* (9 pictures) and *test2* (10 pictures)

	$\overline{A_{occ}}$	$\overline{B}$	V
Test1(9 pictures)	0.1713	0.0081	0.0308
Test2(10 pictures)	0.0412	0.00001	0.0076

5 Conclusion and Future Works

In this paper we have presented a novel approach for photo sticking able to realize an effective image summarization of a set of pictures. Existing approaches require user assistance or change the relative appearance of each picture 9. The original work presented in 12 has been modified by making use of a self-adaptive image cropping algorithm, exploiting visual and semantic information together with an optimization process based on a genetic algorithm. Preliminary results confirm the ability of the proposed method to generate sub-optimal results. The evaluation is mainly based on saliency region occlusion, layer ordering, canvas usage and image cropping. Future works will be devoted to designing further optimization strategies to improve overall robustness also capable of speeding up the overall process. We also plan to introduce compositional rules aimed to cope with the landscape images. Furthermore, we will investigate a quantitative methodology to evaluate collage results by conducting some subjective tests.

References

1. Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S., Colburn, A., Curless, B., Salesin, D., Cohen, M.F.: Interactive digital photomontage. ACM Trans. Graph. 23(3), 294–302 (2004)
2. Atkins, C.B.: Adaptive photo collection page layout. In: ICIP, pp. 2897–2900 (2004)
3. Carson, C., Belongie, S., Greenspan, H., Malik, J.: Blobworld: Image segmentation using expectation-maximization and its application to image querying. IEEE Trans. Pattern Anal. Mach. Intell. 24(8), 1026–1038 (2002)
4. Fei-Fei, L., Fergus, R., Perona, P.: A bayesian approach to unsupervised one-shot learning of object categories. In: ICCV, pp. 1134–1141 (2003)
5. http://fotoinsight.co.uk/book/
6. Chen, J., Chu, W., Kuo, J., Weng, C., Wu, J.: Tiling Slideshow. In: Proceedings of ACM Multimedia Conference, pp. 25–34 (2006)
7. Diakopoulos, N., Essa, I.: Mediating photo collage authoring. In: Proceedings of the 18th Annual ACM Symposium on User interface Software and Technology (2005)
8. Girgensohn, A., Chiu, P.: Stained Glass Photo Collages. In: Proceedings of UIST 2004 Companion, October 24, 2004, pp. 13–14 (2004)
9. Rother, C., Kumar, S., Kolmogorov, V., Blake, A.: Digital tapestry. In: Proceedings of CVPR 2005 Computer Vision and Pattern Recognition, vol. 2005(1), pp. 589–596 (2005)
10. Rother, C., Bordeaux, L., Hamadi, Y., Blake, A.: AutoCollage. ACM Transactions on Graphics 25, 847–852 (2006)

11. Snavely, N., Seitz, S.M., Szeliski, R.: Photo Tourism: Exploring photo collections in 3D. ACM Transactions on Graphics 25(3) (August 2006)
12. Wang, J., Sun, J., Quan, L., Tang, X., Shum, H.Y.: Picture Collage. In: Proceedings of CVPR 2006 – Computer Vision and Pattern Recognition, pp. 347–354 (2006)
13. Ciocca, G., Cusano, C., Gasparini, F., Schettini, R.: Self- Adaptive Image Cropping for Small Displays. In: Proceedings of IEEE ICCE 2007 - International Conference on Consumer Electronics (2007)
14. Chen, L., Xie, X., Fan, X., Ma, W., Zhang, H.J., Zhou, H.Q.: A visual attention model for adapting images on small displays. Multimedia Systems 9, 353–364 (2003)
15. Suh, B., Ling, H., Bederson, B.B., Jacobs, D.W.: Automatic Thumbnail Cropping and its Effectiveness. In: Proc. UIST 2003, pp. 95–104 (2003)
16. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and Regression Trees. Wadsworth and Brooks/Cole (1984)
17. Schettini, R., Brambilla, C., Cusano, C., Ciocca, G.: Automatic classification of digital photographs based on decision forests. IJPRAI 18(5), 819–846 (2004)
18. De Ponti, M., Schettini, R., Brambilla, C., Valsasna, A., Ciocca, G.: Content-Based Digital-Image Classification Method. European Patent no. EP1102180 (2001)
19. Itti, L., Koch, C.: A model of saliency based visual attention of rapid scene analysis. IEEE Trans. on PAMI 20, 1254–1259 (1998)
20. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57, 137–154 (2004)
21. Gasparini, F., Corchs, S., Schettini, R.: Recall or precision oriented strategies for binary classification of skin pixels. Pattern Recognition (submitted, 2006)
22. Haupt, R.L., Haupt, S.E.: Practical Genetic Algorithms, 2nd edn. John Wiley & Sons, Hoboken, New Jersey (2004)