

Saliency Based Aesthetic Cut of Digital Images

Luca Greco and Marco La Cascia

DICGIM, Università degli Studi di Palermo, Italy
{luca.greco,marco.lacascia}@unipa.it

Abstract. Aesthetic cut of photos is a process well known to professional photographers. It consists of cutting the original photo to remove less relevant parts close to the borders leaving in this way the interesting subjects in a position that is perceived by the observer as more pleasant. In this paper we propose a saliency based technique to automatically perform aesthetic cut in images. We use a standard method to estimate the saliency map and propose some post processing on the map to make it more suitable for our scope. We then apply a greedy algorithm to determine the cut (i.e. the most important part of the original image) both in the cases of free and fixed aspect ratio. Experimental results are reported showing how the cut resulting from our technique compares to some state of the art retargeting and cropping techniques.

Keywords: Aesthetic cut, image editing, saliency.

1 Introduction and Related Works

Image saliency refers to those elements that a human observer consider important or attracting his attention. Sometime analyzing the visibility and the most visible elements of an image is also used the term conspicuity.

An inner part of the image is important if it is different from the rest of other regions, gaining greater visibility and attracting attention, resulting in catching the eyes focus. Salient parts of a scene evoke a strong visual response and polarize attention, detaching this zones from background. Human visual attention is composed of two factors connected with two stimuli of different nature:

- the objective factor depends exclusively on the characteristics of the image;
- the subjective factor is related to the subject's will.

The first has a bottom-up activation, and is related to physical characteristics such as brightness, color and shape. The second has a top-down activation and is influenced by the knowledge obtained by learning the probabilistic structure of the environment. In many situations the second give the greatest contribution to the acquisition of information.

Saliency can be used in Computer Vision tasks for smart thumbnailing or subject extraction. In this paper saliency map is used to get an aesthetic cut of pictures.

After shooting professional photographers can improve picture composition, without digital images processing, by simply using editing tools. It is common to crop images for professionals but not for the amateur photographer. Some techniques for image composition, for example Gooch et al. [14] and Byers et al. [11], are based explicitly on the rules of photography and in particular the *rule of thirds*. Santella et al. [10] proposed, in addition, an interactive method for cropping by determining the position of the important content on the basis of the eye tracking. Recently, Liu et al. [4] proposed a method to optimize the composition of the image by applying the basic rules of photographic composition, as the rule of thirds, the predominance of the diagonals, the visual balance, resulting in a score for the aesthetic composition of the image.

2 Saliency

Attention based strategies are helpful in reducing the time of search or analysis. In fact, the analysis of an image is characterized by what you would expect to find or what you want to identify. Attention is composed of two separate processes as claimed by Itti and Koch[12]:

- A bottom-up process, which is based exclusively on a physical stimulus. This process is simple and fast but lacks from any high level considerations.
- A top-down process, which depends on the observer will and is based on a semantic analysis of the image. This process has a higher level of abstraction but it is slower. This search is more accurate having a high degree of environmental knowledge and analyzing the consistency of the context.

In this paper we will deal with bottom-up stimulus and in particular we will use the saliency map model presented in [1].

2.1 Graph-Based Visual Saliency

The GBVS is a model of bottom-up visual saliency, which is composed of two stages: formation of the activation maps of certain features such as color, intensity and brightness, and feature normalization to emphasize the saliency and to allow combination with other maps. First stage consist in extracting feature maps and creating activation maps, constructing a fully-connected graph with a node for each pixel and directed edges weighted using pixels' dissimilarity in feature map and their distance. The weights are normalized to one and used as the transition probability of a Markov chain, then the activation map is calculated by the equilibrium distribution of the resulting chain. Second stage is performed constructing another graph using the calculated activation map and the distances. Again, this graph is used as a Markov chain and the normalized map is derived by the equilibrium distribution.

2.2 Saliency Map

The GBVS saliency map is a matrix containing the saliency of each pixel. The map is normalized, containing values between zero and one indicating the importance of pixels: the closer the saliency is to one, more important is the corresponding pixel. To get an intuition, saliency maps are usually shown as gray scale images (see Figure 1). White pixels correspond to saliency equal to one, black pixels correspond to saliency equal to zero and the shades of gray correspond to the intermediate values. Therefore in correspondence of the two deers of Figure 1 there is a higher concentration of the saliency while in the areas corresponding to the background, i.e. the trees and the terrain, the concentration of saliency is almost uniform.



Fig. 1. Example of an image (a) and the corresponding saliency map (b)

2.3 Noise in the Map

It is worth noting that the estimation of the saliency map, being based on low-level image data, in some cases is not completely correct. For example the presence of cluttered or non-uniform background in the image affects the correct estimation of the saliency map. Also in images containing many subjects in a small area, the saliency map tends to consider the whole area salient without distinguishing the various subjects. Moreover, if a subject has characteristics (shape, color, orientation, etc.) similar to the background it is considered not salient.

3 Subjects Detection

The proposed algorithm is designed to automatically perform an aesthetic cut on images. The algorithm is based on the detection of the most interesting part of the image. This region is then used to determine how to perform the cut. A saliency map is used to determine the area where interesting subjects are present and to separate them from the background.

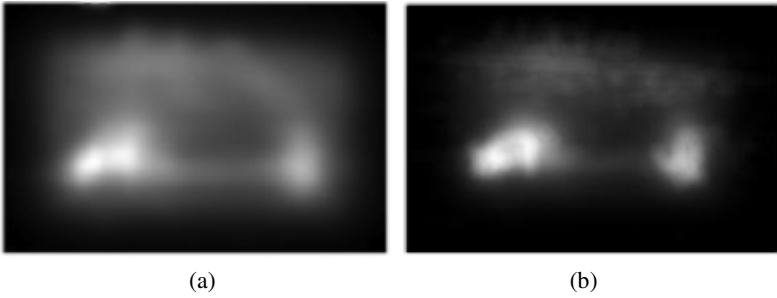


Fig. 2. Example of saliency map before (a) and after (b) the refinement

3.1 Computing and Refining the Saliency Maps

The saliency map used in our work is purely dominated by objective stimulus without taking into account the global image as it is obtained by GBVS algorithm. To perform an aesthetic cut the interesting part of the image is the one associated to subjects while the background can be somewhat left out.

The term “subject” has a different meaning if the image is a close representation of an object or a panoramic view. In the first case subjects are the detail of the represented object, in the second are the most visually attractive locations of the view. In both cases the subjects are what we want to keep in the image after the aesthetic cut. If in the image there are one or more subjects a more useful map is composed of saliency spots corresponding to the subjects. To obtain such a map the application of three filters, each one with a specific goal, is proposed.

The first filter emphasize the saliency in correspondence of image contours to decrease the possibility of an incomplete cut. In fact, if the subject has not a compact shape it may occur that the contour details are ignored in the saliency map. To prevent this, a binary map describing the pixels that belong to the edges is computed using a Sobel filter and a fixed threshold. Then, in the saliency map, the pixels corresponding to edges are incremented by a factor of 0.3.

The second filter is a nonlinear transformation of the saliency map increasing the difference between background pixels and salient pixels. After the transformation less informative pixels are truncated using a fixed threshold. The nonlinear transformation is:

$$y = x^2. \quad (1)$$

Finally, the third filter is simply a Gaussian filter that has the result of smoothing the refined saliency map. Figure 2 shows the result of the filters applied to the saliency map of the image containing the two deers (Figure 1).

3.2 Image Cropping

Based on the refined saliency map we define a cropping algorithm based on a greedy search. A greedy algorithm attempts to construct a pseudo-optimal solution from a partial initial solution and searches for a better solution until a stop criterion is met. The search of the solution does not consider all possible solutions because this could be very expensive. Only some solutions are considered in particular those that are closer to the partial solution, proceeding to the attainment of the pseudo-optimal solution for small steps.

These extensions of a partial solution, that could be called "local", are somewhat smaller and possible extensions are relatively few. Among the various local extensions the algorithm proceeds choosing the most "greedy", or rather the most convenient, which is being promoted new partial solution and is configured as a local optimal solution.

The cropping algorithm based on saliency is initialized with a partial solution obtained by cutting a box centered at the peak of the salience of fixed size (5x5 pixels is the proposed initial box). Then we proceed to increase the dimension of the box obtained at each iteration of partial solutions considering the most valid solutions the one with a higher percentage of salience.

Partial solutions are selected to obtain at each iteration a pane that possesses an increasing saliency. Possible solutions are provided by changing two factors:

1. The center of the frame. The center of the subject does not correspond with the peak, saliency increase is then evaluated by modifying the center of the box along the diagonal directions;
2. The dimensions of the box. In most cases the optimal cropping is not a square, thus evaluating along width or length direction.

3.3 Stopping Criterion

The stopping criterion should determine if the partial frame obtained contains sufficient saliency of the image and therefore if it defines the final cut. At each iteration the sum of internal saliency is calculated and the search algorithm is stopped when it exceeds a threshold.

4 Aesthetic Crop

The threshold used in the stopping criterion defines the dimension of the final cut of the image. Using a high threshold value the result is a frame that contains almost entirely the starting image cutting off only a small part.

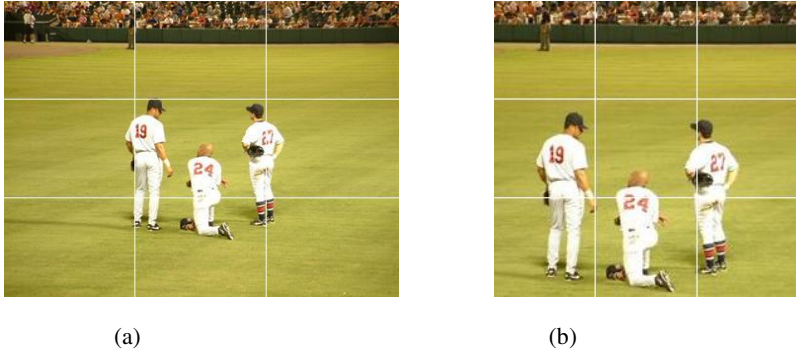


Fig. 3. Original image (a) and cropping result (b). The visual balance of the result is increased, because subjects are distributed over the center and closer to the thirds than in the original image.

4.1 Aesthetics

It is necessary to define an aesthetic measure of an image because there is not a common accepted agreement about what, for example, makes a shot more aesthetic than another. Every photographer and every critic has a personal taste in recognizing aesthetic feature of a picture and in the overall judgment or comparison.

Nevertheless, for educational purposes there are some guidelines mostly accepted and taught in professional or semi-professional courses. Some of these are:

- **Avoid amputations:** subjects of the image have to be entirely included, avoiding to cut off some part of these.
- **Visual Balance:** salient objects have to be equally distributed around the center of the image, so having in it the center of mass of saliency to balance the overall distribution of interesting feature in the picture.
- **Rule of thirds:** the image is divided by two horizontal and two vertical lines equally spaced. The four intersections are the attention points and the subjects or the detail to focus have to be placed in this position. If there are vertical or horizontal elements (like walls or horizon), these have to be placed in the thirds lines.
- **Diagonal dominance:** if diagonal elements are present they have to be located along the diagonals of the frame.

4.2 Proposed Method and Aesthetic Composition Rules

The problem can be formalized as the search for a frame that maximizes the aesthetic of the photo. Therefore taken two crops of the same image we need to define some

measures that can determine which of the crops is aesthetically better. To do this we rely on the saliency map and we consider a crop better than the other if it has a higher percentage of saliency.

Therefore, we present a new cropping algorithm redefining the stopping criterion. For this task, the greedy search of our method stops when the percentage of saliency in the frame is more than a fixed threshold (0.95 in our experiments).

Even though our method does not explicitly take into account composition rules, we observed that resulting crops tend to respect some of those. The experimental observation can be easily explained by analyzing how these rules are related to the saliency in the image.

- **Avoid amputations:** The saliency map considers subjects as salient part of image. Thus, if the saliency map is accurate using a high value of the threshold lead to a complete inclusion of the subjects in the cropped image then avoiding amputations.
- **Visual balance:** In determining the center of the image the algorithm proceeds to change the center of the frame finding the maximum increment of saliency. Using high values of the threshold the center of the crop is not limited to be near the area with the highest concentration of local saliency and it is an approximation of the center of mass of saliency. Figure 3 is an example of a balanced cropped image.
- **Rule of thirds:** as shown in visual balance, the center of the crop isn't constrained to be neither in the center of the original image nor focused on the saliency peak. So, not respecting explicitly this rule, if the original image was not respecting rule of thirds the visual balance of the image after cropping will lead principal subjects to be near the thirds points because salient part of the image are in a zone that is around the center, corresponding to attention points (Figure 3).
- **Predominance of diagonals:** this rule is not respected in the proposed method.

Other methods explicitly apply the aesthetic guidelines, for example segmenting salient objects position or lines in the image. Many works [9][13][4], mostly related to the retargeting problem, use these features to get a more aesthetic image changing salient objects position in the image and not only selecting an internal crop. The result is an aesthetically enhanced picture but not representing the original real scene.

Our method respects the original content and it only selects the internal crop, trying to maximize the aesthetic result. Cropped images are not optimal for retargeting problem.

Figure 4 shows the result of our method and the results of [9][13][4], lines of thirds and attention points are depicted to show how visual balance and rule of thirds are respected in the different methods.

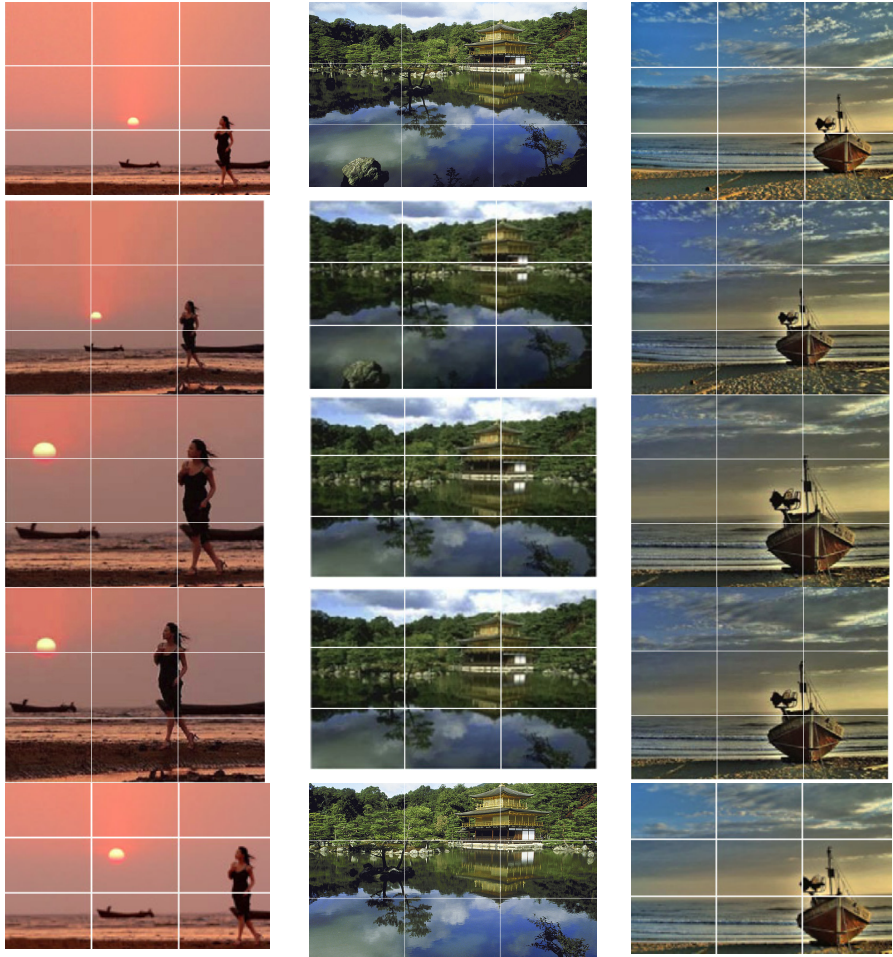


Fig. 4. Result of image aesthetic cut. From first to last row: original image, result of [9], result of [4], result of [13], result of our method.

4.3 Proportional Crop

The proposed greedy algorithm has no constraints regarding the aspect ratio of the resulting cut. This behavior is acceptable only if the crop is used for the personal digital collection or web sharing of images, but is not useful if the pictures have to be printed, because usually there are fixed print dimension.

A possible solution consists of taking the maximal part fitting the desired aspect ratio from the crop resulting from previously described technique. A modified version of the greedy algorithm can directly take into account this constraint estimating at each step the new cropping window respecting the final aspect ratio. In this way no hand-made re-crop is required and the process is totally automatic. Figure 5 show the not constrained and the $4/3$ proportional crop.



Fig. 5. Initial image (a), free cutting result (b), 4/3 constrained cut (c)

5 Conclusions

In this paper we presented a novel saliency-based aesthetic cut technique. The saliency maps used follow a visual bottom-up saliency model and are refined to be more suited for the aesthetic cut problem.

Using the refined map, a greedy algorithm has been used to detect the pseudo-optimal cut including a large amount of the saliency of the map. Even though this is done without explicitly taking into account aesthetic common composition rules we observed that the resulting cuts tend to respect some of these rules.

The results of the proposed technique are shown and visually compared to some state of the art retargeting and cropping techniques. The proposed technique is also able to perform a fixed proportion cut for image to be printed or for specific needs (i.e. web sharing with fixed proportion).

Finally the algorithm has a low computational cost as fine details of composition rules are not directly taken into account but approximately enforced as side effect of the saliency based greedy cut.

References

1. Harel, J., Koch, C., Perona, P.: Graph-Based Visual Saliency. In: Proceedings of Neural Information Processing Systems, NIPS (2006)
2. Joshi, D., Datta, R., Fedorovskaya, E., Luong, Q.-T., Wang, J.Z., Li, J., Luo, J.: Aesthetics and emotions in images, a computational perspective. *IEEE Signal Processing Magazine*, 94–115 (2011)
3. Bhattacharya, S., Sukthankar, R., Shah, M.: A Holistic Approach to Aesthetic Enhancement of Photographs. *ACM Transactions on Multimedia Computing, Communications and Applications* 7S(1), Article 21 (2011)
4. Liu, L., Chen, R., Wolf, L., Cohen-Or, D.: Optimizing Photo Composition. Technical Report (2010)
5. Zhang, M., Zhang, L., Sun, Y., Feng, L., Ma, W.: Auto Cropping for Digital Photographs. Microsoft Research Asia (2005)
6. Avidan, S., Shamir, A.: Seam carving for content-aware image resizing. In: Proceedings. SIGGRAPH Conference, vol. 26 (2007)
7. Suh, B., Ling, H., Bederson, B.B., Jacobs, D.W.: Automatic Thumbnail Cropping and its Effectiveness. Department of Computer Science, University of Maryland (2003)

8. Marchesotti, L., Cifarelli, C., Csurka, G.: A framework for visual saliency detection with applications to image thumbnailing. Xerox Research Centre Europe (XRCE), France (2009)
9. Liu, L., Jin, Y., Wu, Q.: Realtime aesthetic image retargeting. In: Deussen, O., Jepp, P. (eds.) International Symposium on Computational Aesthetics in Graphics, Visualization and Imaging, pp. 1–8. Eurographics Association, London (2010)
10. Santella, A., Agrawala, M., Decarlo, D., Salesin, D.H., Cohen, M.: Gaze-based interaction for semiautomatic photo cropping. In: ACM Human Factors in Computing Systems (CHI), pp. 771–780. ACM Press (2006)
11. Byers, Z., Dixon, M., Smart, W.D., Grimm, C.: Say cheese! experiences with a robot photographer. pp. 25–37–46 (2004)
12. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* (1998)
13. Jin, Y., Wun, Q., Liu, L.: Aesthetic photo composition by optimal crop-and-warp. Department of Mathematics, Zhejiang University
14. Gooch, B., Reinhard, E., Moulding, C., Shirley, P.: Artistic composition for image creation. In: Eurographics Workshop on Rendering Technique, pp. 83–88. Eurographics Association (2011)
15. Suh, B., Ling, H., Bederson, B.B., Jacobs, D.W.: Automatic Thumbnail Cropping and its Effectiveness. Department of Computer Science, University of Maryland
16. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.-Y.: Learning to detect a salient object. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(2), 353–367 (2011)