A complexity approach for identifying aesthetic composite landscapes

Adrian Carballal, Rebeca Perez, Antonino Santos, and Luz Castro

Department of Information and Communication Technologies
University of A Coruña, A Coruña, Spain
{adrian.carballal,rebeca.perezf,nino,maria.luz.castro}@udc.es

Abstract. The present paper describes a series of features related to complexity which may allow to estimate the complexity of an image as a whole, of all the elements integrating it and of those which are its focus of attention. Using a neural network to create a classifier based on those features an accuracy over 85% in an aesthetic composition binary classification task is achieved. The obtained network seems to be useful for the purpose of assessing the Aesthetic Composition of landscapes. It could be used as part of a media device for facilitating the creation of images or videos with a more professional aesthetic composition.

1 Introduction

An evaluator which enable users without an artistic background to take pictures of better appearance could be used by different multimedia devices. It could help users to identify in real time those framings with a certain aesthetic value. It could be also used for different tasks related to aesthetic composition: identification, classification, categorization, etc.; both in real-time multimedia devices and in stand-alone applications.

Unlike those systems which allow access to the images intrinsic information based on different image phenomena, such as contrast, determine the aesthetic composition of a given image is not a trivial task.

Numerous papers [9, 16, 11] have appeared in recent years evaluating different elements of the aesthetic value of images and different ways to estimate it. This paper introduces the uses of different metrics based on those works, based on the complexity of an image, which have already proven useful in experiments related to the ordering and classification based on stylistic and aesthetic criteria [15, 8, 13, 14]. These metrics and their usefulness for calculating the aesthetic composition of a landscape is studied.

The present paper is structured as follows: i) a short description of the state of the art in composition systems is included; ii) a set of complexity metrics and their usefulness for calculating the aesthetic composition of a landscape is presented and the features to be used in the study are described; iii) the results obtained in an experiment of image classification according to their aesthetic composition are shown; iv) the overview of a photography expert about the results is shown; v) and, finally, the conclusions and the upcoming research lines for improving the already presented approach will be explained.

2 State of the Art

Previous works focus on the search for metrics which show the composition quality or cropping methods which enhance the visual and aesthetic quality of a given image. Most of them use metrics related to Rule of Thirds (RoT) [5], Region of Interest (ROI) [21], or Saliency [17] individually. RoT is a photographic framing technique which divides the scene into 9 equally sized parts by means of three vertical and horizontal equidistant lines. This technique is based on placing the heaviest elements at the intersection among these lines. On the other hand, the use of ROI determines those image areas grouping the elements which attract the greatest interest. Saliency detection allows the differentiation of a foreground object from the background and to classify it as an interesting point.

Santella et al. [17] presented a system which records user's eye movements for a few seconds to identify important image content. The given approach is capable of generating crops of any size or aspect ratio. The main disadvantage is that the system incurs on requiring user input, so it can't be considered a fully-computational approach. Once the important area of an image is detected, the crops are made considering three basics on photography: (i) include an entire subject and some context around, (ii) edges should pass through featureless areas whenever possible, (iii) the area of the subject matter should be maximized to increase clarity. They presented 50 images cropped using 3 different approaches: saliency-based [18], professional hand-crop, and gaze-crop to 8 different subjects. They obtained that their gaze-based approach was preferred to saliency-based cropping in 58.4% of trials and in 32.5% to professional cropping.

Liu et al. [5] have translated several basic composition guidelines into quantitative aesthetic scores, including the Rule of Thirds, diagonal, visual balance, and region size. Based on which, an automatic crop-and-retarget approach to producing a maximally beautiful version of the input image. Their approach searches for the optimal composition result in a 4D space which contains all cropped windows with various widths and heights. A dataset of 900 casual images arbitrarily collected from international websites in which skilled photographers rank photographs through them was employed to evaluate their score function. To evaluate the performance of their method generated a set of 30 triplets of images; the original image, one crop using Santella's method and one using theirs. These triplets were shown to 56 subjects, males and women, between 21 and 55 years old. In 44.1% of cases, the subjects preferred the cropped images provided by their approach. In addition, 81.8% were not able to distinguish whether the image was hand-cropped or computationally optimized.

Wang and Cohen [19] propose an algorithm for composing foreground elements onto a new background by integrating matting and compositing into a single process. The system is able to compose more efficiently and with fewer artifacts compared with previous approaches. The matte is optimized in a sense that it will minimize the visual artifacts on the final composed image, although it may not be the true matte for the foreground. They determine the size and position that minimizes the difference between a small shell around the foreground and the new background, and then run the compositional matting. The developed

algorithm not always gives satisfying compositions when the new background differs significantly from the original.

Zhang et al. [21] presented an auto-cropping model to obtain an optimal cropped image using the width and height of the original image, the conservative coefficient, the faces detected and the Region of Interest (ROI). The model consists of three sub models: (i) a composition sub model to describe how good the composition is, (ii) a conservative sub model to prevent the photograph from being cropped too aggressively and (iii) a penalty factor to prevent faces or ROIs being cut off. They used 100 pictures randomly selected from 600 home photographs. All the images were used into two studies. The first user study evaluated the auto cropping result in different aspect ratios. They obtained that the algorithm exhibits a satisfactory score on cropping. The second user study evaluated the improvement of the picture composition after cropping, in which observed the considering of the artistic rules leads to a good score of the improvement of the picture composition.

Suh et al. [18] proposed a set of fully automated image cropping techniques using a visual salience model based on low-level contrast measures. According to them, the more salient a portion of image, the more informative it is; and the visual search performance is increased as much recognizable the thumbnail is. They used their feature set on recognizing objects in small thumbnails (Recognition Task) and to measure how the thumbnail generation technique affects search performance (Visual Search Task). They ran an empirical study over 20 subjects, which were college or graduate students at the University of Maryland, and 500 filler images. In both tasks, the proposed set was capable to provide thumbnails substantially more recognizable and easier to find in the context of visual search.

What if the quality of aesthetic composition may be related to the visual complexity of the composition itself, as well as to the complexity derived from each of the elements represented in the same image? We assume that, inside the images, there are elements which attract the observer's attention and their complexity must be taken into account when determining the composition aesthetics. This assumption takes into account what Philip Galanter calls "effective complexity" [3]. According to him, an artist will tire both high ordered and high disordered aesthetic composition because of lack any structural complexity worth. We propose the joint use of metrics which allow to estimate the complexity of an image as a whole, as well as of all the elements integrating it and, particularly, those which are its focus of attention. Following this hypothesis, this work will centre concretely on the study of the aesthetic composition of landscapes. The proposed metrics are listed next.

3 Complexity metrics and presented features

Machado and Cardoso [9], based on previous works by [1], proposes JPEG and FRACTAL compression methods to estimate the image complexity. On [2] it is found a correlation between compression error and complexity of the image.

The error involved in the JPEG compression method, which affects mainly to high frequencies, depends on the variability of the pixels in the image. From this point of view, more variability involves more randomness and therefore more complexity. The FRACTAL method tends to compress an image by filtering the self-similarities within. In this case, more self-similarities implies less variability, and therefore less complexity. Hence we considered applying JPEG and FRACTAL Compression methods as image complexity estimatives.

Since both are lossy compression schemes, there might be a compression error, i.e., the compressed image will not exactly match the original. Three levels of detail for the JPEG and FRACTAL compression metrics are considered: low, medium and high. More info available at [14].

Salience is the quality that stands out one or multiple important objects from those that surrounds it/them. Somehow, saliency facilitates to focus the perception of the viewer on the most pertinent item or items on a scene. The saliency algorithm chosen to implement was the Subject Saliency algorithm also known as Subject Region Extraction [6]. Based on the idea that the subject in a photograph would be clearer and the background would be blurred, the algorithm extracts the clear region of an image which theoretically holds the subject. This algorithm uses images statistics to detect 2D blurred regions in an image, based on a modification of [4] work. Subject Salience will be used to detect the foreground item/s, which should get the focus of attention.

The Sobel Filter calculates the gradient of the image intensity at each point, giving the direction of the greater variation from light to dark and the amount of variation in that same direction. This gives us an idea of the variation of brightness at each point, from smooth to sharp differences. With this filter it is estimated the presence of the light-dark transitions and how they are oriented. With these light-dark variations corresponding to the intense and well-defined boundaries between objects, it is possible to obtain edge detection. The Sobel Filter will give a simple representation of all the elements standing on the image by identifying their silhouettes.

We have also chosen to use a set of basic features related to the statistical variability of the pixels integrating an image. Said features calculate: (i) the mean (ii) and the standard deviation of the pixels with regard to the adjacent pixels in each color channel.

It must be explained the way in which our features set will be obtained. Four auxiliary images are generated from every image. Three of those images are obtained by separating the color channels following the HSV model. The fourth image stems from an attempt to solve the existing problems of the HSV color model for the extreme values of the H and V channels. When pixel value of $V = 0$ the resulting color is always black. In the case of $S = 0$ then the resulting color is always in grayscale. Another problem occurs when $V = 255$ and $S = 0$, in which the color is white. In any of these situations it can not be assured that data from H are correct as it depends directly from the transformation algorithm to HSV format. In an attempt to address this deficiency, a new image is determined by

multiplying pixel by pixel the S and V channels within the range [0, 255]. It will be referred to from now on as CS or Colorfulness [12].

Splitting the image in mentioned color channels and applying the complexity metrics to each of the resulting images gives a total of 24 features (from now feature set COMPLEX). It must be noted that these 24 features will be calculated based on the original image, having applied the Subject Saliency and the Sobel Filter again. Therefore, we will achieve a total number of 72 features (set COMPLEX^F). Since Hue channel is circular, the mean and standard deviation are calculated based on the angle values of Hue and its norm. In addition, it is performed the multiplication of the Hue angle by the pixel intensity values of CS, and a new value of the norm is calculated using values from H and CS. It yields another 12 features per image (from now set BASE), 7 related with mean (set AVG) and 5 related with standard deviation (set STD), giving another set of 36 features per image if also applying Subject Saliency and Sobel Filter (set BASE^F).

4 Experimental dataset for aesthetic composition

A total of 1961 landscape images of high aesthetic quality in their composition have been compiled for carrying out this experiment, most of them wallpapers in landscape format. All of them have a resolution higher than 1024x1024 pixels. Their visual topics vary a great deal: night, day, mountain, beach, etc.

From this initial dataset, a random algorithm was created which will provide sub-images with a width/height ratio equal to the original image (see functioning in Algorithm 1). Said algorithm has been used on every image, thus providing a second set of images of the same sampling size.

A photography expert has identified those images which, because of the random cropping, generated a new image which was equal or better than the original one as regards framing. All these images have been discarded, achieving a final dataset integrated by two sets of 1757 images each.

Figure 1 shows a simple subset of images of both sets. The left side shows images of the original set, while the right side shows the same subset once the algorithm has been applied.

5 Classification method and results obtained

Both images to be used in this experimental part and the features which will characterize them individually have been presented so far. The present section explains the experiment carried out.

A classification model has been developed using SNNS (Stuttgart Neural Network Simulator) [20]. In particular, a backpropagation MLP is used with 3-layer architecture: an input layer with 108 neurons, a single hidden layer of 15 neurons and an output layer with 1 neuron. This configuration has been established based on previous experiments and experiences of the research team in tasks of the same field [10, 7].

Algorithm 1: Random Image Cropping
1 for *each image* **do**
2   A random height is established (between 400px and 1/2 of the height of the original image)
3   A width is established according to the ratio of the original image
4   Random locx and locy are created (< 0, $>$ original size)
5   **if** *the cropped image does not exceed the original on the right or at the bottom:* **then**
6   accept cropping
7   **else**
8   return to 2

The network training will finish when a maximum number of 1000 cycles is reached. The initial network weights are determined at random within the range [-0.1, 0.1]. A maximum error tolerance of 0.3 has been used. The 10-fold Cross-Validation (10-fold CV) model has been used for the generation of the training data sets so that their results are statistically relevant. Each of these runs has a different training and validation set which have been randomly generated. The results shown correspond to the average results obtained in these 10 runs.

Given that the neural network provides a number value within the range of 0 and 1, a binary system has been used for cataloguing the images. Those images which have a network output, once they have been presented to the system, of less than 0.5 will be catalogued as having a low aesthetic composition. The classification results given for each feature set presented in Section 3 are shown in Table 1.

Table 1: Precision and Recall using ANNs

Feature Set	#feats	Accuracy	**RECALL**		**PRECISION**	
			Cropped	Original	Cropped	Original
COMPLEX[F]	72	85.74%	82.64%	88.84%	88.11%	83.65%
COMPLEX	24	80.88%	76.49%	85.26%	83.84%	78.39%
BASE[F]	36	78.66%	73.99%	83.32%	81.61%	76.21%
BASE	12	75.67%	69.04%	82.30%	79.59%	72.66%
AVG	7	71.60%	68.70%	74.50%	72.93%	70.41%
STD	5	74.67%	67.79%	81.56%	78.61%	71.69%

According to the data, it may be seen that the image classification when using the BASE set seems to achieve relatively satisfactory results. It should be noted that the problem itself contributes to the achievement of such high results. Let's imagine that there is a landscape photography similar to the one in Figure

(a) Original image 1

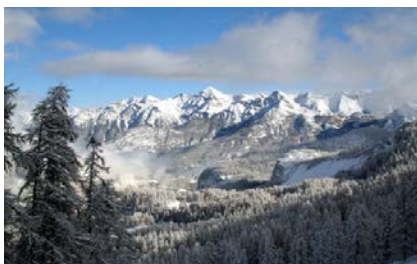(b) Cropped image 1'

(c) Original image 2

(d) Cropped image 2'

(e) Original image 3

(f) Cropped image 3'

(g) Original image 4

(h) Cropped image 4'

Fig. 1: Example images of both sets (images of the original set on the left and the cropped version on the right)

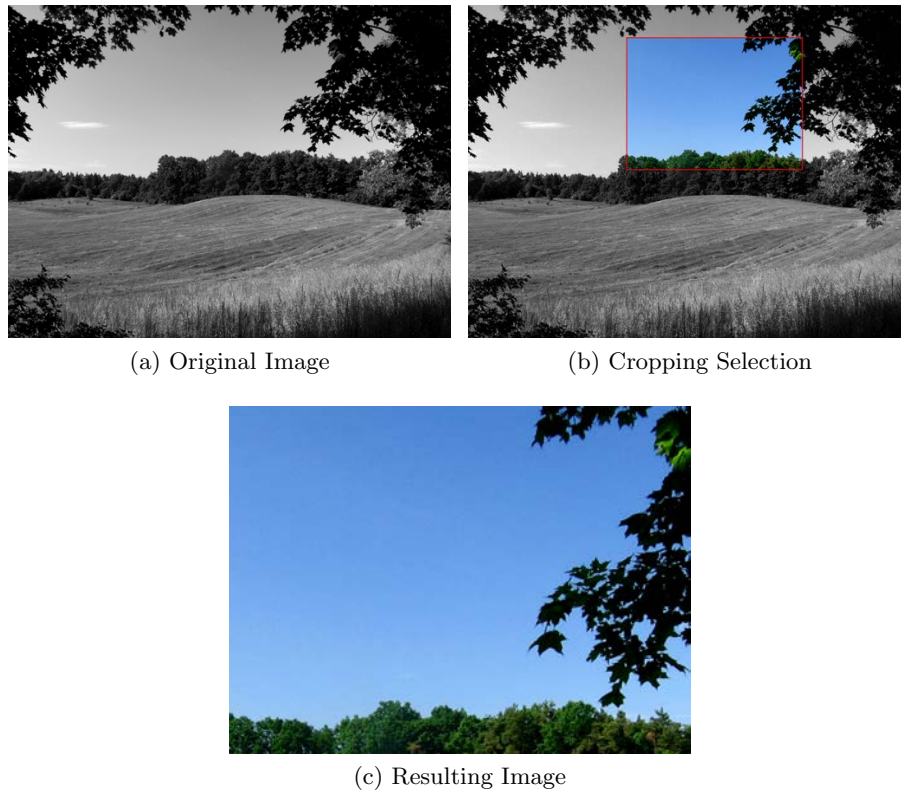2a. Having applied the random cropping, the new image may result as the one seen in Figure 2c.

(a) Original Image

(b) Cropping Selection

(c) Resulting Image

Fig. 2: Cropping example

In this case, as is usually the case in this kind of images, the cropped landscapes usually have an extreme pixel variability compared to the original image. That is, the mean and the standard deviation of the pixels integrating the resulting image either increases or decreases considerably. Anyhow, no element of the real content of any of the two images is taken into account for the classification.

The network trained with COMPLEX^F set offers a success rate of 85.74%. Individually, the network succeeds 88.84% of the original set versus 82.64% of cropped one. Considering precision, we obtained a rate of 83.66% relative to originals and 88.11% relative to cropped images. The results seem to be well-balanced globally, both in accuracy and precision.

The fact of not using edge detection filters (COMPLEX set) derives in a decrease in both recall and precision, which demonstrates that to attend to the important elements present on a landscape provides relevant information. On

the other hand, if we compare with BASE^F is seen as a lesser set of complexity metrics provide better results than the filters applied on different measures of variability of the pixels.

The observed results seem to indicate that to attend to the complexity of the important elements inside a landscape allow to differentiate those images that present a better composition.

5.1 An expert overview

The mentioned expert was presented with the images which the classifier was and wasn't capable of identifying correctly, without having any kind of information about the method used.

According to the expert, the classifier seems to work correctly in those not obvious cases where there is some element of brightness, light or where the differentiating element is relatively small with regard to the image (Figure 3a). Even on those images whose content bears a great symmetry, regardless of their content or originality (Figure 3b).

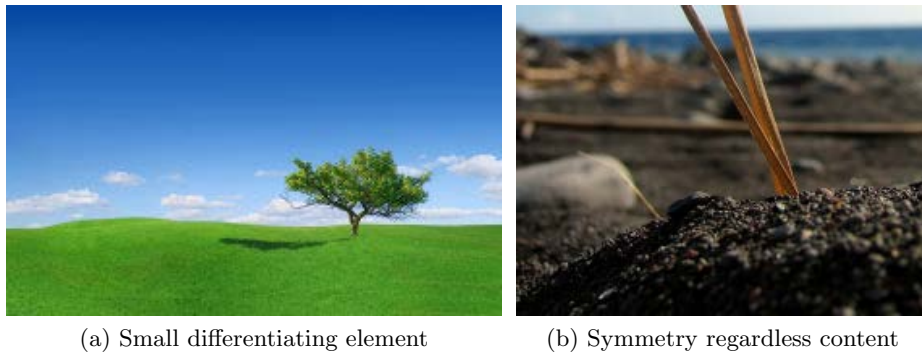(a) Small differentiating element   (b) Symmetry regardless content

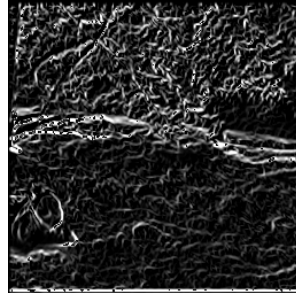Fig. 3: Examples of well classified images.

In case of false positives, one of the most frequent mistakes happens when the system cannot find any differentiating element in the image (Figure 4a). The resulting image after applying the Sobel Filter seems to contain a lot of noise compared to those images where there is a clear differentiator that the edge detector identifies correctly (Figure 4b).

On the contrary, false negative cases, sometimes the mistake is perfectly justified: cropped image pieces partially comply with the principles of framing; they structure a differentiating element at the center while their environment goes totally unnoticed as a uniform background (Figure 5).

The expert concluded that most of the mistakes generated by the presented classification method were not trivial and were sometimes understandable.

(a) Applying no filter   (b) Applying Sobel Filter

Fig. 4: Example of a wrongly classified image because of the Sobel Filter.

(a) Original   (b) Cropped

Fig. 5: Example of a wrongly classified image as original being cropped

6 Future work

This paper has presented a set of metrics based on complexity which seem to be useful for judging the aesthetic composition in landscape images. A neural network has been used as a binary classifier using the presented features as inputs, achieving accuracy and precision results over 85%.

This classifier could be add to any real-time multimedia device and used for different tasks related to aesthetic composition like identification or classification, categorization, etc. For this propose we intend to modify the classification system so that asynchronous tasks can be performed by means of parallel programming, thus reducing the time of the task of extracting metrics from each image, which currently entails the biggest bottleneck.

Among the most immediate enhancements, we may mention above all the elimination of all those cases identified by the expert where the classifier fails, both in the case of false positives and negatives. For this purpose, we intend to search for another set of metrics which can help the already existing one with that task, and even to find alternatives for Sobel and Saliency Subjects, so that their detection problems do not have a direct impact on the prototype.

7 Acknowledgments

References

1. Arnheim, R.: Art and Visual Perception, a Psychology of the Creative Eye. Faber and Faber, London (1956)
2. Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C.J., Sawey, M.: Predicting beauty: Fractal dimension and visual complexity in art. British Journal of Psychology 102(1), 49–70 (2011)
3. Galanter, P.: What is generative art? complexity theory as a context for art theory. In: International Conference on Generative Art. Milan, Italy (2003)
4. Levin, G., Feinberg, J., Curtis, C.: Alphabet synthesis machine. http://alphabet.tmema.org (2006)
5. Liu, L., Chen, R., Wolf, L., Cohen-Or, D.: Optimizing photo composition. Comput. Graph. Forum 29(2), 469–478 (2010)
6. Luo, Y., Tang, X.: Photo and video quality evaluation: Focusing on the subject. In: Forsyth, D.A., Torr, P.H.S., Zisserman, A. (eds.) ECCV (3). Lecture Notes in Computer Science, vol. 5304, pp. 386–399. Springer (2008)
7. Machado, P., Cardoso, A.: All the truth about NEvAr. Applied Intelligence, Special Issue on Creative Systems 16(2), 101–119 (2002)
8. Machado, P., Romero, J., Cardoso, A., Santos, A.: Partially interactive evolutionary artists. New Generation Computing – Special Issue on Interactive Evolutionary Computation 23(42), 143–155 (2005)
9. Machado, P., Cardoso, A.: Computing aesthetics. In: Oliveira, F. (ed.) Proceedings of the XIVth Brazilian Symposium on Artificial Intelligence: Advances in Artificial Intelligence. LNCS, vol. 1515, pp. 219–229. Springer, Porto Alegre, Brazil (1998)
10. Machado, P., Romero, J., Manaris, B.: Experiments in computational aesthetics: An iterative approach to stylistic change in evolutionary art. In: Romero, J., Machado, P. (eds.) The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music, pp. 381–415. Springer Berlin Heidelberg (2007)
11. Rigau, J., Feixas, M., Sbert, M.: Informational dialogue with van gogh's paintings. In: Eurographics Symposium on Computational Aesthetics in Graphics, Visualization and Imaging. pp. 115–122 (Jun 2008)
12. Romero, J., Machado, P., Carballal, A., Correia, J.: Computing aesthetics with image judgement systems. In: McCormack, J., dInverno, M. (eds.) Computers and Creativity, pp. 295–322. Springer Berlin Heidelberg (2012), http://dx.doi.org/10.1007/978-3-642-31727-9_11
13. Romero, J., Machado, P., Carballal, A., Osorio, O.: Aesthetic classification and sorting based on image compression. In: Chio, C.D., Brabazon, A., Caro, G.A.D., Drechsler, R., Farooq, M., Grahl, J., Greenfield, G., Prins, C., Romero, J., Squillero, G., Tarantino, E., Tettamanzi, A., Urquhart, N., Etaner-Uyar, A.S. (eds.) EvoApplications (2). Lecture Notes in Computer Science, vol. 6625, pp. 394–403. Springer (2011)
14. Romero, J., Machado, P., Carballal, A., Santos, A.: Using complexity estimates in aesthetic image classification. Journal of Mathematics and the Arts 6(2-3), 125–136 (2012)

15. Romero, J., Machado, P., Santos, A., Cardoso, A.: On the development of critics in evolutionary computation artists. In: Günther, R., et al. (eds.) Applications of Evolutionary Computing, EvoWorkshops 2003: EvoBIO, EvoCOMNET, EvoHOT, EvoIASP, EvoMUSART, EvoSTOC. LNCS, vol. 2611. Springer, Essex, UK (2003)

16. Ross, B.J., Ralph, W., Hai, Z.: Evolutionary image synthesis using a model of aesthetics. In: Yen, G.G., Lucas, S.M., Fogel, G., Kendall, G., Salomon, R., Zhang, B.T., Coello, C.A.C., Runarsson, T.P. (eds.) Proceedings of the 2006 IEEE Congress on Evolutionary Computation. pp. 1087–1094. IEEE Press, Vancouver, BC, Canada (16–21 Jul 2006)

17. Santella, A., Agrawala, M., DeCarlo, D., Salesin, D., Cohen, M.: Gaze-based interaction for semi-automatic photo cropping. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 771–780. CHI '06, ACM, New York, NY, USA (2006)

18. Suh, B., Ling, H., Bederson, B.B., Jacobs, D.W.: Automatic thumbnail cropping and its effectiveness. In: UIST. pp. 95–104. ACM (2003)

19. Wang, J., Cohen, M.F.: Simultaneous matting and compositing. In: Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007). IEEE Computer Society (2007)

20. Zell, A., Mamier, G., Vogt, M., Mache, N., Hübner, R., Döring, S., Herrmann, K.U., Soyez, T., Schmalzl, M., Sommer, T., et al.: SNNS: Stuttgart Neural Network Simulator User Manual, version 4.2. Tech. Rep. 3/92, University of Stuttgart, Stuttgart (2003)

21. Zhang, M., Zhang, L., Sun, Y., Feng, L., Ma, W.Y.: Auto cropping for digital photographs. In: ICME. pp. 438–441 (2005)