Learning Visual Balance from Large-scale Datasets of Aesthetically Highly Rated Images

Ali Jahanian[a], S. V. N. Vishwanathan[b], and Jan P. Allebach[a]

[a]School of Electrical and Computer Engineering;
[b]Departments of Computer Science and Statistics;
Purdue University, West Lafayette, IN 47907, U.S.A

ABSTRACT

The concept of visual balance is innate for humans, and influences how we perceive visual aesthetics and cognize harmony. Although visual balance is a vital principle of design and taught in schools of designs, it is barely quantified. On the other hand, with emergence of automantic/semi-automatic visual designs for self-publishing, learning visual balance and computationally modeling it, may escalate aesthetics of such designs. In this paper, we present how questing for understanding visual balance inspired us to revisit one of the well-known theories in visual arts, the so called theory of "visual rightness", elucidated by Arnheim. We define Arnheim's hypothesis as a design mining problem with the goal of learning visual balance from work of professionals. We collected a dataset of 120K images that are aesthetically highly rated, from a professional photography website. We then computed factors that contribute to visual balance based on the notion of visual saliency. We fitted a mixture of Gaussians to the saliency maps of the images, and obtained the hotspots of the images. Our inferred Gaussians align with Arnheim's hotspots, and confirm his theory. Moreover, the results support the viability of the center of mass, symmetry, as well as the Rule of Thirds in our dataset.

Keywords: Visual balance, Arnheim's theory of visual rightness, layout, aesthetics, automatic visual design, the Rule of Thirds, symmetry, design mining.

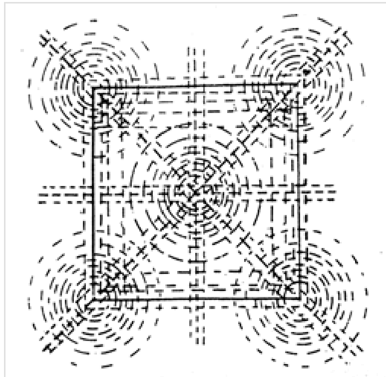
1. INTRODUCTION

Psychological studies show that visual balance is an innate concept for humans [1, 2], which influences how we perceive visual aesthetics and cognize harmony [3]. There exists a body of work endeavouring to understand visual balance and its relation with symmetry [4] about vertical [5–7] and horizontal [7] axes, content of the scene [8], color contrast [9, 10], and styles in abstract and representational artworks [11–14].

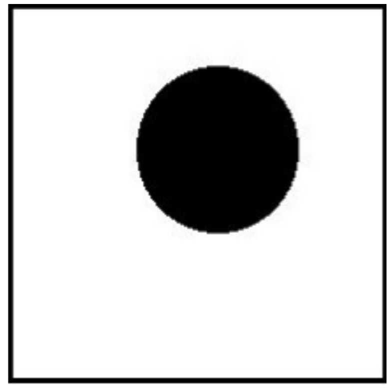
In visual design, for instance, balance is a key principle that helps designers to convey their messages [15–19]. Photographers, specifically, create visual balance in the spatial composition of photos through photo cropping [20, 21]. Our motivation in this work is to model visual balance. Learning visual balance from the work of professionals in design and photography may help to enable the automatic design applications in layout creation [22–29], content retargeting [30–32], cropping [21, 33], photo composition [32, 34] and quantifying aesthetics of layouts [35–40]. Nevertheless, there is no rigorous model to describe visual balance. In prior studies, the references to this notion are mainly based on art theorists' speculations and general guidelines from professional designers. However, because of access to large-scale datasets, we might be able to revisit such a theoretical concept in art and attempt for a more quantifiable definition.

In prior work, visual balance is defined as "looking visually right" [41] and is studied under the "theory of rightness in composition" [11, 20]. Balance is considered in two general categories: symmetry and asymmetry [42], which in any case relates to harmony [43]. One of the central theories around balance is perhaps Arnheim's structural net [44] (see Fig. 1 (a)), in which he hypothesizes that there are nine hotspots (including the center) on any visual artwork, and identifies their locations. Although in prior work Arnheim's net is studied through psychophysical experiments, by asking participants' opinions about spatial arrangements of visual elements in paintings and photos, to the best of our knowledge, the present work is the first design mining framework on large-scale datasets of images for evaluating Arnheim's theory.

Further author information: (Send correspondence to Ali Jahanian)
Ali Jahanian: E-mail: ali-design@csail.mit.edu, Telephone: 1 765 464 9030



(a)



(b)

Figure 1. Arnheim's structural net. This net illustrates the hotspots on any visual artwork. b) In this figure, Arnheim explains that when the disk is not centered in the canvas, there are visual forces to our eyes that attempt to drag the disk to the center. Images reproduced from [44].

In this paper, we examine Arnheim's hypothesis about balance through design mining a dataset of 120K images. These images are obtained from a professional photography website, 500px [45], and have at least 100 user likes (some of these images have several thousands of likes). Because visual balance is stimulated by the interaction of visual elements such as lines, color, texture, and orientation in an image, we run our design mining on the images' saliency maps. This decision is justified according to the fact that a similar set of visual elements comprises the underlying features in the saliency map models. Having computed the saliency maps of the images, we then model these maps as a mixture of Gaussians by utilizing GMM and EM techniques. In our modeling, we initially position the Gaussians in the same places that Arnheim locates his visual balance hotspots on an image. We describe the adaption of GMM and the scalability considerations in processing large-scale datasets in our framework. Our inferred Gaussians align with Arnheim's hotspots, and confirm his structural net, specifically the center and the symmetry of the net.

The flow of this paper is as follows. In Sec. 2, we first discuss the theories behind visual balance, and how we intuitively model it. We then describe our formal definition and modeling framework in Sec. 3. The results of our work are presented in Sec. 4. We conclude this paper in Sec. 5 with a general discussion and future work.

2. BACKGROUND

2.1 Visual Balance in Spatial Composition

Visual balance is often studied along with the spatial composition of an image. Spatial composition in an image is defined as the arrangement of visual elements and the way that they interact with each other in the space. Perhaps one of the earliest attempts in quantifying spatial structure of visually appealing artworks belongs to Adolf Zeising through defining *Golden Section* ratios [46]. Gustav Fechner later studied the Golden Section ratios, and argued that this notion is overemphasized [47]. Some studies argue against the Golden Section concept (e.g. [48]), and some are in favor of it (e.g. [49]). This could be because of the way that the experiments were set up in Fechner's studies [50]. Another well-known rule in creating visually appealing spatial composition is called the *Rule of Thirds*. While some argue in favor of this rule, others discuss its triviality [51]. The notion of visual balance is another component of the spatial composition that remains a challenge. In studies of balance to this date, this concept has almost always been equated with the physical connotation of balance and equilibrium (e.g., see [20]). However, visual balance to art experts might metaphorically mean harmony [43]. This inspires us to study this notion beyond a measure of weight along the vertical or horizontal axes. We aim to revisit Arnheim's structural net, and examine it through mining from a large-scale dataset of highly liked images. We attempt to answer the following question: In this large-scale dataset, can we infer any pattern to support Arnheim's speculation about his hypothetical hotspots in his proposed structural net? In short, our modeling framework indeed supports Arnheim's net.

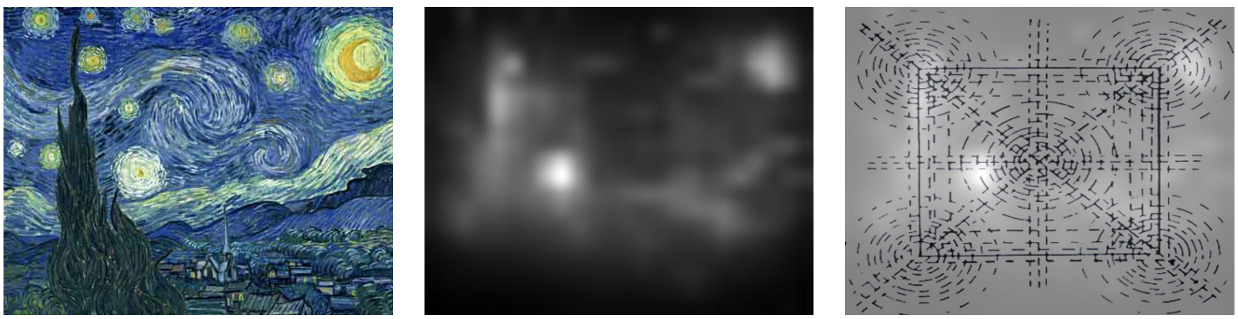
2.2 Theory of Visual Rightness

As mentioned earlier, visual balance is defined as "looking visually right" [41] and is studied under the "theory of rightness in composition" [11, 20] in prior work. One of the central theories around balance is perhaps Arnheim's structural net [44].

Figure 1 (a) illustrates Arnheim's net, in which he hypothesizes that there are nine hotspots (including the center) on any visual artwork, and identifies their locations. When visual weights are located on these hotspost, visual stability and balance are more perceived. For instance, see Fig. 1. In this figure, Arnheim explains that because the disk is not positioned at the center of the canvas, there are visual forces to our eyes attempting to drag the disk to the center. This is because we prefer the state of equilibrium. These forces have visual weight and direction.

According to Arnheim, factors that influence visual weight are: dynamic, position, depth, size, color, intrinsic interest, isolation, shape simplicity, shape orientation, and knowledge of the scene [44]. Some of these factors are studied in prior work, e.g. features that influence dynamic quality of static abstract designs [52].

Because a similar set of visual elements contributes to visual saliency or conspicuity, we model Arnheim's visual weight with visual saliency. We justify this decision by the following example. Figure 2 illustrates an image (The Starry Night by Vincent van Gogh) in (a), its saliency map in (b) computed by the algorithm described in [53], and in (c) the overlap of the saliency map of the image with Arnheim's structural net (Fig. 1 (a)). As observed, in the overlap image the salient areas of the original image are almost aligned with the hotspots in the Arnheim's net. This observation inspired us to revisit Arnheim's net and seek whether such an observation can be validated in a large number of professional artworks (photographs in our dataset). In Sec. 3, we suggest a formal definition for our observation.



(a) (b) (c)

Figure 2. An example for overlap of saliency map with Arnheim's net. a) The original image, The Starry Night by Vincent van Gogh. b) The saliency map of the original image computed by using the algorithm described in [53]. Note that in this saliency map, the more salient pixels are visualized by the lighter pixels. c) The overlap of (b) with Arnheim's net (in Fig. 1 (a)). It is observed that the salient areas of the original artwork are almost aligned with the hotspots of Arnheim's net.

3. MODELING FRAMEWORK

A saliency map (e.g. Fig. 2 (b) of an image is a grayscale image with pixels corresponding to the pixels in the original image. Each saliency pixel can be represented by a vector $(x, y, v)^\top$, where x and y correspond to the spatial location of the pixel, and $v \in 0, 1, ..., 255$, the luminance value of the pixel, corresponds to the saliency value of this pixel. In other words, a higher value of v represents a more salient pixel.

Our goal is to model the saliency values of a saliency map by a mixture of Gaussians. Therefore, we represent a saliency map as a scatterplot of points in the 2D Cartesian space, and fit a Gaussian mixture model to the density distributions of this scatterplot. We first define our Gaussian mixture and then, in the next section, compute this mixture by the Expectation Maximization (EM) algorithm.

First, we define a saliency scatterplot of a saliency map as follows. For each pixel $(x, y, v)^\top$ in the saliency map, we generate v numbers of point $\mathbf{x} = (x, y)^\top$ in its corresponding saliency scatterplot. We denote a saliency scatterplot with N number of points in set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$. In this fashion, we can represent the value of a saliency pixel as a measure of density of points in the corresponding saliency scatterplot. For the Gaussian mixture analysis, we follow the notations in [54]. The Gaussian mixture distribution for point $\mathbf{x}$ in its saliency scatterplot can be written as a linear combination of Gaussians in the form

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{1}$$

where the mixing coefficient π_k is modeled as the probability of assigning pixel $\mathbf{x}$ to Gaussian component k. To model the mixing coefficients, we introduce a K dimensional binary random variable $\mathbf{z}$ drawn from a multinomial distribution, where $p(z_k = 1) = \pi_k$. Hence, this distribution can be written in the form

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}. \tag{2}$$

Because π_k is a probability, it must satisfy $0 \leqslant \pi_k \leqslant 1$ together with

$$\sum_{k=1}^{K} \pi_k = 1. \tag{3}$$

Intuitively, the advantage of defining the mixing coefficients as probabilities is that we can introduce a K dimensional binary random variable $\mathbf{z} = (z_1, z_2, ..., z_K)^\top$, where dimension z_k is the label or assignment of point $\mathbf{x}$ to Gaussian component k with some probability π_k. Variable $\mathbf{z}$ is latent to us, and by defining a joint distribution of it and observed point $\mathbf{x}$, we can infer the assignments using the Bayes formula. Formally, we define the joint distribution of $\mathbf{x}$ and $\mathbf{z}$ as

$$p(\mathbf{x}, \mathbf{z}) = p(\mathbf{x} \mid \mathbf{z}) \cdot p(\mathbf{z}). \tag{4}$$

Note that we can compute $p(\mathbf{x})$ by marginalizing this joint probability over $\mathbf{z}$. Because $\mathbf{z}$ is latent, we can infer it based on the observation of point $\mathbf{x}$ using the Bayes formula

$$\gamma(z_k) \equiv p(z_k = 1 \mid \mathbf{x}) = \frac{p(z_k = 1)\, p(\mathbf{x} \mid z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)\, p(\mathbf{x} \mid z_j = 1)} = \frac{\pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}\left(\mathbf{x} \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)}. \tag{5}$$

In fact, in our modeling, π_k are priors for $z_k = 1$, and $\gamma(z_k)$ is the corresponding posterior after observation of $\mathbf{x}$.

So far, we have defined the problem for one point. Our problem for our dataset of saliency scatterplots is as follows. Denote $\mathcal{X}$ as our dataset of I number of saliency scatterplots. We assume there are K Gaussian components underlying in each scatterplot, with mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$; however, each component has a different mixing coefficient π_{ik} for the i-th saliency scatterplot and the Gaussian mixture component k. To fit a Gaussian mixture to a saliency scatterplot in dataset $\mathcal{X}$, we denote the i-th saliency scatterplot by $\mathbf{X}^{(i)}$, for $i = 1, 2, \ldots, I$, and represent it as an $N \times D$ matrix, where N denotes the number of the points in the i-th saliency scatterplot, and $D = 2$ denotes the dimensions of each point. Note that N for each saliency scatterplot is different; however, we do not index it for simplifying the notations. In this notation, the n-th row in matrix $\mathbf{X}$ is point $\mathbf{x}_n$ given by $\mathbf{X}_n^\top$. If we assume that the points in a saliency scatterplot are independent, then the log of the likelihood function for all the I saliency scatterplots stored in $\mathcal{X}$ is given by

$$\ln p(\mathcal{X} \mid \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{I} \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_{ik} \mathcal{N}\left(\mathbf{x}_n^{(i)} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \right\}. \tag{6}$$

By maximizing the log likelihood function in (6) with respect to the means $\boldsymbol{\mu}_k$, we obtain

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{I} \sum_{n=1}^{N} \gamma^{(i)}(z_{nk})\, \mathbf{x}_n^{(i)}, \tag{7}$$

where

$$N_k = \sum_{i=1}^{I} \sum_{n=1}^{N} \gamma^{(i)}(z_{nk}),$$

and

$$\gamma^{(i)}\left(z_{nk}\right) = \frac{\pi_{ik}\mathcal{N}\left(\mathbf{x}_n^{(i)} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}{\sum_{j=1}^{K} \pi_{ij}\mathcal{N}\left(\mathbf{x}_n^{(i)} \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)}. \tag{8}$$

By maximizing the log likelihood function with respect to the covariance matrixes $\boldsymbol{\Sigma}_k$, we obtain

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k}\sum_{i=1}^{I}\sum_{n=1}^{N}\gamma^{(i)}\left(z_{nk}\right)\left(\mathbf{x}_n^{(i)} - \boldsymbol{\mu}_k\right)\left(\mathbf{x}_n^{(i)} - \boldsymbol{\mu}_k\right)^{\top}. \tag{9}$$

Finally, we obtain the mixing coefficients π_{ik} for saliency scatterplot $\mathbf{X}^{(i)}$, by maximizing the log likelihood function with respect to π_{ik} while taking account of the constraint (3) using a Lagrange multiplier:

$$\pi_{ik} = \frac{N_k^{(i)}}{N}, \tag{10}$$

where

$$N_k^{(i)} = \sum_{n=1}^{N}\gamma^{(i)}\left(z_{nk}\right). \tag{11}$$

3.1 EM for Gaussian Mixtures

Given the Gaussian mixture model for our dataset, we optimize to maximize the likelihood function in (6) with respect to its parameters $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$, and π_{ik}. The EM algorithm is then:

1. Initialize $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$, and π_{ik}, and compute the initial value of the log likelihood.
2. **E step.** Compute $\gamma^{(i)}\left(z_{nk}\right)$.
3. **M step.** Recompute the estimation of the parameters using the current $\gamma^{(i)}\left(z_{nk}\right)$.
4. Compute the log likelihood again, and check for the convergence condition. If the condition is satisfied, stop; otherwise return to step 2.

3.2 Dataset

Our dataset includes about 120K images. These images are obtained from a professional photography website, 500px [45], and have at least 100 user likes (some of these images have several thousands of likes). These images also have semantic tags (e.g. *landscape*, *architecture*, *fashion*, etc.) which enable us, for future work, to cluster some general visual balance templates and to establish their linkage to the tags. The preprocessing of the images is performed with Matlab Imaging and Parallel Computing toolboxes. This includes image resizing and computing of the saliency maps.

4. RESULTS

We performed our algorithm * on the Purdue Clusters using Matlab Parallel Computing toolbox. We assigned 96 cores in 24 nodes for the parallel computing. For our experiment, we consumed about 120 hours for the EM algorithm. We initiated two types of Gaussians for our experiment, five Gaussians as illustrated in Fig. 3 (a) and nine Gaussians in Fig. 3 (c). We initiated these Gaussians to text Arnheim's hotspots (compare with Fig. 1 (a)).

The results of our computations are illustrated in Fig. 3 (b) and Fig. 3 (d) for initial Gaussians in Fig. 3 (a) and Fig. 3 (c), respectively. Our computed Gaussian mixtures align with Arnheim's hotspots, and confirm his structural net, specifically the center and the symmetry of the net.

---

*Our implementation is available upon request.

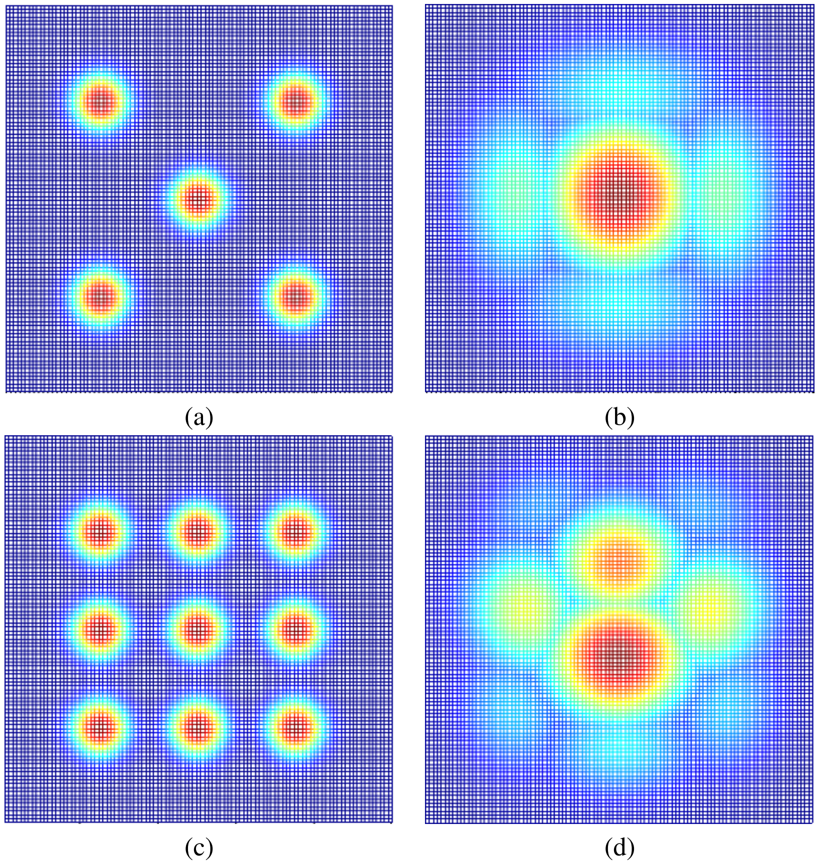


Figure 3. Initial and fitted Gaussian mixtures computed by our framework. In (a) and (c) five and nine (respectively) initial Gaussians defined in the mixture to resemble the hotspots at the corners and the center of Arnheim's net. b) and d) show the fitted Gaussians initialized in (a) and (c), respectively.

Moreover, the second column of Fig. 3 suggests that Arnheim's power of the center [55], or the center of the mass is the most important location in the images. This supports the idea of computing the center of the mass as a means of quantifying balance. It is also observed that the Rule of Thirds is a viable notion in the images (especially in Fig. 3 (b)). However, this needs more investigation. For instance, one may argue that our images are mainly taken by experts or photographers who respect such a rule of thumb.

5. DISCUSSION AND FUTURE WORK

In this paper, we discussed some of the theoretical aspects of the concept of visual balance in images and artworks. We argued that this concept needs to be revisited through design mining large-scale datasets. For this account, we gathered a dataset of 120K highly liked images obtained from a professional photography website, 500px [45]. We developed a computational framework to model important or salient parts of the images with a mixture of Gaussians. Our goal was to examine what Arnheim had speculated about the existence of stable axes and hotspots in an image. Arnheim suggested that the overlap of visual weights with these hotspots may represent a feeling of balance. The fitted Gaussian mixtures by our framework align with Arnheim's hotspots, and support his structural net. The results specifically confirm the center and the symmetry of the net.

At this stage, our analysis supports Arnheim's structural net. However, we believe that further investigation is necessary to understand how experts and non-experts prefer such a structure in images. Similar to some of the experiments in the recent work of McManus et al. [20], we need to study random or low-liked images as well. One valid question is whether the photographers of highly liked images were aware of or even trained in some of the rules for positioning important elements in certain locations. Another question is whether by distorting the saliency structure of images we can still obtain acceptable visually balanced compositions.

Another part of our future work is to study semantics of visual balance. Because our dataset contains semantic tags (e.g. *landscape*, *architecture*, *fashion*, etc.) of the images, for future work, we aim to cluster some general visual balance templates, and to establish their linkage to the tags. This may lead to recommendations for visual layout of photos and visual design in general. For instance, in photography, it might be intuitive to layout the horizon line on one of the horizontal axes of Arnheim's net. As another example, in automatic design of magazine covers, assuming there is a human face in the cover image, we might be able to recommend a set of good candidate places to layout the face. A similar recommendation may be made for locating a brand logo in a design.

ACKNOWLEDGMENTS

We gratefully thank David Sigman, Head of The Patti and Rusty Rueff School of Visual and Performing Arts at Purdue University for his input on the concept of visual balance.

References

[1] Bornstein, M. H., Ferdinandsen, K., and Gross, C. G., "Perception of symmetry in infancy," *Developmental Psychology* **17**(1), 82 (1981).

[2] Giannouli, V., "Visual symmetry perception," *Encephalos* **50**, 31–42 (2013).

[3] Palmer, S. E., Schloss, K. B., and Sammartino, J., "Visual aesthetics and human preference," *Annual review of psychology* **64**, 77–107 (2013).

[4] Norcia, A. M., Candy, T. R., Pettet, M. W., Vildavski, V. Y., and Tyler, C. W., "Temporal dynamics of the human response to symmetry," *Journal of Vision* **2**(2), 1 (2002).

[5] Wilson, A. and Chatterjee, A., "The assessment of preference for balance: Introducing a new test," *Empirical Studies of the Arts* **23**(2), 165–180 (2005).

[6] Sammartino, J. and Palmer, S. E., "Aesthetic issues in spatial composition: Effects of vertical position and perspective on framing single objects.," *Journal of Experimental Psychology: Human Perception and Performance* **38**(4), 865 (2012).

[7] Gershoni, S. and Hochstein, S., "Measuring pictorial balance perception at first glance using japanese calligraphy," *i-Perception* **2**(6), 508 (2011).

[8] Leyssen, M. H., Linsen, S., Sammartino, J., and Palmer, S. E., "Aesthetic preference for spatial composition in multiobject pictures," *i-Perception* **3**(1), 25 (2012).

[9] Linnett, C. M., Morriss, R. H., Dunlap, W. P., and Fritchie, C. J., "Differences in color balance depending upon mode of comparison," *The Journal of general psychology* **118**(3), 271–283 (1991).

[10] Polat, U. and Tyler, C. W., "What pattern the eye sees best," *Vision Research* **39**(5), 887–895 (1999).

[11] Locher, P. J., "An empirical investigation of the visual rightness theory of picture perception," *Acta Psychologica* **114**(2), 147–164 (2003).

[12] Locher, P., Overbeeke, K., and Stappers, P. J., "Spatial balance of color triads in the abstract art of Piet Mondrian," *Perception-London* **34**(2), 169–190 (2005).

[13] Vartanian, O., Martindale, C., Podsiadlo, J., Overbay, S., and Borkum, J., "The link between composition and balance in masterworks vs. paintings of lower artistic quality," *British Journal of Psychology* **96**(4), 493–503 (2005).

[14] Tyler, C. W., "Some principles of spatial organization in art," *Spatial Vision* **20**(6), 509–530 (2007).

[15] Locher, P. J., Jan Stappers, P., and Overbeeke, K., "The role of balance as an organizing design principle underlying adults' compositional strategies for creating visual displays," *Acta Psychologica* **99**(2), 141–161 (1998).

[16] White, A. W., [*The elements of graphic design: space, unity, page architecture, and type*], Skyhorse Publishing Inc. (2002).

[17] Cullen, K., [*Layout workbook: a real-world guide to building pages in graphic design*], Rockport Publishers (2007).

[18] Samara, T., [*Design evolution: theory into practice: a handbook of basic design principles applied in contemporary design*], Rockport Publishers (2008).

[19] Bergström, B., [*Essentials of visual communication*], Laurence King Publishers (2009).

[20] McManus, I., Stöver, K., and Kim, D., "Arnheim's gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images," *i-Perception* **2**(6), 615 (2011).

[21] Santella, A., Agrawala, M., DeCarlo, D., Salesin, D., and Cohen, M., "Gaze-based interaction for semi-automatic photo cropping," in [*Proceedings of the SIGCHI conference on Human Factors in computing systems*], 771–780, ACM (2006).

[22] Jacobs, C., Li, W., Schrier, E., Bargeron, D., and Salesin, D., "Adaptive grid-based document layout," *ACM transactions on graphics (TOG)* **22**(3), 838–847 (2003).

[23] Hurst, N., Li, W., and Marriott, K., "Review of automatic document formatting," in [*Proceedings of the 9th ACM Symposium on Document Engineering (DocEng'09)*], 99–108 (2009).

[24] Merrell, P., Schkufza, E., Li, Z., Agrawala, M., and Koltun, V., "Interactive furniture layout using interior design guidelines," in [*ACM Transactions on Graphics (TOG)*], **30**(4), 87, ACM (2011).

[25] Nebeling, M., Matulic, F., Streit, L., and Norrie, M. C., "Adaptive layout template for effective web content presentation in large-screen contexts," in [*Proceedings of the 11th ACM symposium on Document engineering*], 219–228, ACM (2011).

[26] Jahanian, A., Liu, J., Tretter, D., Lin, Q., Damera-Venkata, N., O'Brien-Strain, E., Lee, S., Fan, J., and Allebach, J., "Automatic design of magazine covers," in [*IS&T/SPIE Electronic Imaging*], 83020N–83020N–8, International Society for Optics and Photonics (2012).

[27] Kuhna, M., Kivelä, I.-M., and Oittinen, P., "Semi-automated magazine layout using content-based image features," in [*Proceedings of the 20th ACM international conference on Multimedia*], *MM '12*, 379–388, ACM (2012).

[28] Jahanian, A., Liu, J., Lin, Q., Tretter, D., O'Brien-Strain, E., Lee, S. C., Lyons, N., and Allebach, J., "Recommendation system for automatic design of magazine covers," in [*Proceedings of the 2013 international conference on Intelligent user interfaces*], 95–106, ACM (2013).

[29] O'Donovan, P., Agarwala, A., and Hertzmann, A., "Learning layouts for single-page graphic designs," (2014).

[30] Setlur, V., Takagi, S., Raskar, R., Gleicher, M., and Gooch, B., "Automatic image retargeting," in [*Proceedings of the 4th international conference on Mobile and ubiquitous multimedia*], 59–68, ACM (2005).

[31] Kumar, R., Kim, J., and Klemmer, S. R., "Automatic retargeting of web page content," in [*CHI'09 Extended Abstracts on Human Factors in Computing Systems*], 4237–4242, ACM (2009).

[32] Liu, L., Chen, R., Wolf, L., and Cohen-Or, D., "Optimizing photo composition," in [*Computer Graphics Forum*], **29**(2), 469–478, Wiley Online Library (2010).

[33] Stentiford, F., "Attention based auto image cropping," in [*The 5th International Conference on Computer Vision Systems, Bielefeld*], Citeseer (2007).

[34] Banerjee, S. and Evans, B. L., "In-camera automation of photographic composition rules," *Image Processing, IEEE Transactions on* **16**(7), 1807–1820 (2007).

[35] Ngo, D. C. L., Samsudin, A., and Abdullah, R., "Aesthetic measures for assessing graphic screens," *J. Inf. Sci. Eng* **16**(1), 97–116 (2000).

[36] Datta, R., Joshi, D., Li, J., and Wang, J., "Studying aesthetics in photographic images using a computational approach," *Computer Vision–ECCV 2006* , 288–301 (2006).

[37] Moshagen, M. and Thielsch, M. T., "Facets of visual aesthetics," *International Journal of Human-Computer Studies* **68**(10), 689–709 (2010).

[38] Lai, C.-Y., Chen, P.-H., Shih, S.-W., Liu, Y., and Hong, J.-S., "Computational models and experimental investigations of effects of balance and symmetry on the aesthetics of text-overlaid images," *International journal of human-computer studies* **68**(1), 41–56 (2010).

[39] Obrador, P., Schmidt-Hackenberg, L., and Oliver, N., "The role of image composition in image aesthetics," in [*Image Processing (ICIP), 2010 17th IEEE International Conference on*], 3185–3188, IEEE (2010).

[40] Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., and Gajos, K. Z., "Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness," in [*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*], 2049–2058, ACM (2013).

[41] Carpenter, P., "Art and ideas: An approach to art appreciation," (1971).

[42] Lok, S., Feiner, S., and Ngai, G., "Evaluation of visual balance for automated layout," in [*Proceedings of the 9th international conference on Intelligent user interfaces*], 101–108, ACM (2004).

[43] Samuel, F. and Kerzel, D., "Judging whether it is aesthetic: Does equilibrium compensate for the lack of symmetry?," *i-Perception* **4**(1), 57 (2013).

[44] Arnheim, R., [*Art and visual perception: A psychology of the creative eye*], Univ. of California Press (1954).

[45] "500px." `http://500px.com/`.

[46] Padovan, R., [*Proportion: science, philosophy, architecture*], Taylor & Francis (1999).

[47] Fechner, G. T., "Various attempts to establish a basic form of beauty: Experimental aesthetics, golden section, and square," *Empirical studies of the arts* **15**(2), 115–130 (1997).

[48] Höge, H., "The golden section hypothesis—its last funeral," *Empirical Studies of the Arts* **15**(2), 233–255 (1997).

[49] Konecni, V. J., "The golden section: Elusive, but detectable," *Creativity Research Journal* **15**(2-3), 267–275 (2003).

[50] McManus, I., Cook, R., and Hunt, A., "Beyond the golden section and normative aesthetics: why do individuals differ so much in their aesthetic preferences for rectangles?," *Psychology of Aesthetics, Creativity, and the Arts* **4**(2), 113 (2010).

[51] Amirshahi, S. A., Hayn-Leichsenring, G. U., Denzler, J., and Redies, C., "Evaluating the rule of thirds in photographs and paintings," *Art & Perception* **2**(1-2), 163–182 (2014).

[52] Locher, P. J. and Stappers, P. J., "Factors contributing to the implicit dynamic quality of static abstract designs," *PERCEPTION-LONDON-* **31**(9), 1093–1108 (2002).

[53] Harel, J., Koch, C., and Perona, P., "Graph-based visual saliency," in [*Advances in Neural Information Processing Systems 19*], 545–552, MIT Press (2007).

[54] Bishop, C. M. et al., [*Pattern recognition and machine learning*], vol. 1, springer New York (2006).

[55] Arnheim, R., [*The power of the center: A study of composition in the visual arts*], Univ. of California Press (1983).