Artistic Composition for Image Creation

Bruce Gooch   Erik Reinhard   Chris Moulding   Peter Shirley
University of Utah

Abstract. Altering the viewing parameters of a 3D object results in computer graphics images of varying quality. One aspect of image quality is the *composition* of the image. While the esthetic properties of an image are subjective, some heuristics used by artists to create images can be approximated quantitatively. We present an algorithm based on heuristic compositional rules for finding the format, viewpoint, and layout for an image of a 3D object. Our system computes viewing parameters automatically or allows a user to explicitly manipulate them.

1 Introduction

Composition is taught to artists by showing them a few simple rules, then showing them a number of pitfalls to avoid. We apply rules from the artistic community as well as observations from the psychology literature. Perhaps it would be more systematic to extract compositional principles entirely from the psychology literature, but what is currently known in that field [15, 18, 21] is not yet specific enough to allow automation. While automation is not needed by artists who know both how to apply and when to break these rules, our system is intended for the more common non-artistic user.

Little work dealing with artistic composition has been published in the computer graphics literature. Feiner and Seligmann [9, 17] borrowed principles from technical illustration. Kawai et al. [11] automated the creation of pleasing lighting. Both He et al. [20] and Karp and Feiner [10] examined how animation sequences are developed. Kowalski et al. [12] have explored user guided composition.

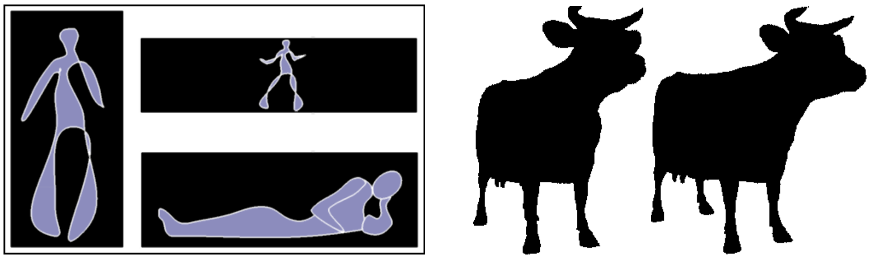
2 Compositional Principles

In art, heuristics for creating images of 3D objects fall into three general categories: choosing the **format** (image size, shape, and orientation); choosing the **viewpoint**; and choosing the **layout** of the object on the image plane.

2.1 Format

The format of an image describes its shape and proportions. An image that is wider than it is tall has a *landscape* format, images that are taller than wide have a *portrait* format. Artists use the following rule of thumb [5], landscape formats should be used with horizontal objects, and portrait formats with vertical objects as in Figure 5 This allows the object to become part of the format rather than dividing it as shown in Figure 1(a).

While the proportions of the format are chosen at the whim of the artist, most art instructors agree that the format of an image should be established first [5]. Early work in psychology showed that the *golden ratio* seems to be preferred [3, 16]. The golden ratio is $(\sqrt{5}+1)/2 \approx 1.618$. Artists often use a five by eight format, which is regarded as being derived from the golden ratio.



(a) The image on the left has a vertical format in accord with the subject. Likewise, in the horizontal lower image. The subject in the upper right image is out of relationship with the format and divides the image.

(b) Left: an "accidental" view where one of the cows hind legs ends up directly behind a front leg. Right: the same cow from a slightly perturbed viewing direction.

Fig. 1. Examples of some formating and viewpoint heuristics.

2.2 Viewpoint

Psychologists have studied viewers' preferences for one viewpoint over another for particular objects. A viewpoint that is preferred by most viewers is called a *canonical viewpoint*. Palmer et al. [13] found that canonical viewpoints are off-axis, while Verfaillie [19] discovered that a three-quarter view of a familiar object is preferred.

A thorough investigation of canonical views was recently carried out by Blantz et al. [6]. They found three predictors of whether a view is canonical: the significance of visible features for a given observer, the stability of the view with respect to small transformations, and the extent to which features are occluded.

Significant features for an observer may include the facial portion of a head, the handle of a tool, or the seat of a chair. In viewing objects, Blantz et al. found that people preferred views which expressed the manner in which an object was seen in its environment, i.e. chairs are viewed from above while airplanes may be viewed from above or below. They also found a distinct lack of "handedness" when humans choose preferred views. For example, when viewing a teapot a right handed viewer did not mind if the handle was placed on the left side of the image.

Image stability means that the viewpoint can be moved with little or no change in the resulting image. Many psychology researchers have shown that objects in a scene which share an edge will confuse a viewer [4, 5, 15]. For example the viewpoint that produces the "three legged cow" in Figure 1(b) is never picked as a canonical view.

When subjects in the Blantz et al. study were given the ability to choose the viewpoint for an object, it was discovered that the subjects performed an internal optimization to find a viewpoint that showed the smallest number of occlusions. This occurred for both familiar objects and artificial geometric constructs. For instance, when choosing a viewpoint for a teapot the subjects always choose a viewpoint that shows both the handle and the spout. This result agrees with Edelman et al. [8] who showed that canonical views for "nonsense" objects may also exist.

Artists have their own heuristics for choosing view directions that are consistent with the psychology results: pick an off-axis view from a natural eye height. Direct $45°$ angles are avoided. Another rule is to have the projections of front/side/top of the object to have relative areas of 4/2/1 on the canvas [2, 18] (often expressed as 55%/30%/15%). The front and side dimensions can be exchanged depending on the object.

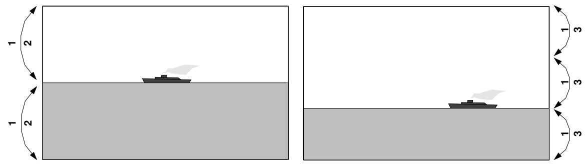


Fig. 2. *Halving the canvas creates static compositions which are peaceful and quiet, but may seem dull. Dividing the canvas into thirds yields a more dynamic image. Note that the rules are applied both horizontally and vertically (after Clifton [7]).*

2.3 Layout

The best known rule of layout is the *rule of thirds* (Figure 2). By partitioning their canvas into thirds both vertically and horizontally, and placing the strong vertical and horizontal components of the image near these lines, artists avoid equal spatial divisions of their image. Equal spatial divisions give an image balance and symmetry. However, equal divisions may also cause an image to be dull, due to the lack of any dynamic quality in the image. Artists have also found the *rule of fifths* useful. Division into quarters is to be avoided because the centerline introduces too much symmetry [7]. These rules can be mixed by dividing the canvas into thirds along one axis and fifths along the other, as in Figure 5.

There are additional, often contradictory, minor layout heuristics taught to artists which are quantifiable. Art theorists contend that the most important information in the image should be placed near the center [3, 18]. However, studies show that objects in a scene should be repelled from the corners and center of the format [2]. Having chosen a viewpoint, it is good practice to place the object in the bottom portion of the image if the viewpoint is above the object or to place the object in the top portion of the image if the viewpoint is below the object. Strong diagonal lines yield a more dynamic image. However, lines oriented toward corners tend to draw the viewers eye off of the image [7].

3 Computer Graphics Implementation

The previous section shows a method for constructing images by first choosing format based on object aspect-ratio. Then choosing the viewpoint to be both off-axis and "natural" for the object. Finally, the object is "framed" within the boundaries of the format to produce a pleasing layout. These steps lead directly to our algorithm.

Our algorithm attempts to find a good composition for a computer graphics image of a 3D object. The algorithm can be run in a fully automatic mode as long as "front" and "top" are defined for the object, but user intervention can be applied at any stage.

We first have the user select a format of either portrait or landscape for a five by eight canvas. Our default is landscape. The format could be found automatically using the principle direction of the orthographic projection of the object. We then compute an initial off axis viewpoint for the object. Finally, we use a robust optimization procedure to perturb the viewing parameters guided by heuristic rules for layout.

3.1 Viewing Parameterization

Of the many possible ways to specify viewing parameters, we choose a system with dimensions that are as intuitive as possible to help us gain insight into the optimization space. We fix two parameters to reduce the dimension of the space we search during the optimization process. The view-up vector is fixed to be parallel to the "top" direction of the model. We also fix the horizontal and vertical field-of-view parameters.

Our free variables are the two spherical coordinates of the vector from the object center to the camera, the two spherical angles of camera pan and tilt relative to that vector, and the distance of the camera to the object center. This gives five free variables, the first two corresponding to rotating position around the object, the second two controlling camera orientation relative to the object, and the last allowing the camera to move toward or away from the object.

3.2 Initial Viewpoint

As a default we choose a viewpoint above and in front of the object. We set left and right arbitrarily due to the finding of Blantz et al. [6] that viewers do not seem to have a preference for left versus right views. The specific three quarter view of the object is set according to the 4/2/1 rule described in Section 2.2. Given the octant the viewpoint resides in there is a unique direction corresponding to the proportions of the orthographic projection of the objects bounding box. Once the initial view direction is fixed, the initial distance from object center to viewpoint is set to be twice the width of the bounding box so we are certain our viewpoint is on screen. Otherwise our layout optimization could converge to a degenerate local minimum created by a blank screen.

3.3 Layout Optimization

Once we have an initial viewing direction, we would like to use a rule such as the rule of thirds, to perturb the viewing parameters into a "good" composition. We would like to detect important image features such as silhouettes, crease lines, strong illumination gradients, and important semantic features like faces. However, we have made our exploratory work as simple as possible and focus only on silhouettes. We would like our optimization procedure to move silhouette lines near third or fifth lines.

We assume that our model is polygonal, with at least a medium level of tessellation, and compute silhouettes in a brute force fashion. If the model occludes a silhouette edge we call that edge a hidden silhouette. For simplicity we do not eliminate hidden silhouettes, and use the silhouette midpoints for computation. We project each midpoints onto a target image with pixel values between zero and one (Figure 6). The target image contains a template with dark pixels near "magnet" features, and light pixels elsewhere. Minor layout heuristics can be combined with the rules of thirds or fifths by compositing their respective templates. Note that any grey scale image could be used to drive our optimization. Figure 5 shows a compositional template inspired by the famous "diamond" composition of Van Gogh's *Irises (1890)*. The objective function is the sum of the pixel values hit by silhouette midpoints. A set of silhouettes that lands mostly on dark pixels is "good", and a set that hits mostly light pixels is "bad". If a midpoint lands off-screen, it takes on the value one plus a linear distance term. This allows edges to be off screen, but encourages them to move toward the screen.

The objective function is reasonably well behaved, although with unknown gradient. This makes the downhill simplex (Nelder-Mead) [14] method well-suited because it does not require analytic derivatives for the objective function.

A concern is that the global minimum for our objective function is to move the camera far away with a pan and tilt that projects all edges onto the darkest pixel. Fortunately, there seem to be enough appealing local minima for this not to occur in practice. Our goal is a reasonable image, instead of the global minimum for the objective function, therefore a local minimum meets our needs. Another concern is that by using midpoints of segments, both short and long edges have equal weight. We could weight edges by length, but equal weighting gives extra importance to highly polygonalized regions which often correspond to preferred semantic features such as faces.

Once the layout optimization has converged, we run a secondary optimization that attempts to eliminate accidental views that arise for coincident silhouettes. A result of this secondary process is shown in Figure 1(b), where the cows hind leg becomes unoccluded. Changing the viewing distance, pan, and tilt do not affect accidental views. Therefore we fix these values and allow the secondary optimization to operate in the two dimensional space of view angles. The objective function that is minimized for this step is one over a constant term plus the sum of squared distances between all midpoints. The constant term keeps the function finite. Although this computation is quadratic on the number of silhouette edges, the objective function is only two dimensional and thus this stage is not a bottleneck. Because we are only trying to climb away from local minima where silhouette edges line up we run the secondary optimization for just 100 iterations.

3.4 Results

Our system was implemented in C on a 250MHz R10000 SGI Origin. Figure 7 shows the results of our algorithm on a 69473 triangle model of a bunny. This image converged in 272 iterations and took approximately three minutes in the initial stage of optimization. The secondary optimization to remove a possible accidental view took a few seconds. Figure 5 shows a 6272 polygon toy plane, with overlaid layout solutions from two initial viewpoints, one above and one below. The solution converged in 165 iterations and took approximately six seconds.

Figure 5 shows the initial viewpoint computed for a 5804 polygon cow model, along with three different layout solutions overlaid on their templates. The rotated template was inspired by the famous “diamond” composition of Van Gogh’s *Irises (1890)*. This image layout converged in 133 iterations and took about five seconds to compute.

4 Conclusions and Future Work

We presented an overview of compositional principles and a proof-of-concept implementation that automates creation of simple images based on quantitative compositional heuristics. There are many directions to take this work. Our objective function operates on silhouette edges which may not correspond to important image features.

Our algorithms work with single objects rather than scenes. In scenes, the grouping of objects should be done in a manner which tells a story about the objects or describes their relationship with one another. There are compositional rules that can serve as guidelines in this process [4, 5, 15]. Calahan [1] explains how lighting can be used to control perceived grouping of scene elements. These processes are highly dependent on scene semantics and may thus be difficult to automate. Advanced composition will most likely remain the domain of the trained artist. However, the increasing number of computer users with no formal artistic training provides a large market for tools that assist in the aesthetic process.

5 Acknowledgments

We would like to thank Brian Smits and Don Nelson for their help in the initial phases of this work. This work was carried out under NSF grants NSF/STC for computer graphics EIA 8920219, NSF/ACR, NSF/MRI and by the DOE AVTC/VIEWS.

References

1. APODACA, A. A., AND GRITZ, L. *Advanced Renderman Creating CGI for Motion Pictures*. Morgan Kaufmann, 2000.
2. ARNHEIM, R. *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press, 1974.
3. ARNHEIM, R. *The Power of the Center*. University of California Press, 1988.
4. BARBOUR, C. G., AND MEYER, G. W. Visual cues and pictorial limitations in photorealistic images. *The Visual Computer 9*, 4 (1992), 151–165.
5. BETHERS, R. *Composition in Pictures*. Pitman Publishing Corporation, 1964.
6. BLANZ, V., TARR, M. J., AND BULTHOFF, H. H. What object attributes determine canonical views. *Perception 28*, 5 (1999), 575–600.
7. CLIFTON, J. *The Eye of the Artist*. North Light Publishers., 1973.
8. EDELMAN, S., AND BULTHOFF, H. Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research 32*, 12 (1992), 2385–2400.
9. FEINER, S. Apex: an experiment in the automated creation of pictorial explanations. *IEEE Computer Graphics & Applications 5*, 11 (November 1985), 29–37.
10. KARP, P., AND FEINER, S. Issues in the automated generation of animated presentations. In *Graphics Interface* (1990), pp. 39–48.
11. KAWAI, J. K., PAINTER, J. S., AND COHEN, M. F. Radioptimization - goal based rendering. In *Proceedings of SIGGRAPH* (1993), pp. 147–154.
12. KOWALSKI, M. A., HUGHES, J. F., RUBIN, C. B., AND OHYA, J. User-guided composition effects for art-based rendering. *2001 ACM Symposium on Interactive 3D Graphics* (March 2001), 99–102. ISBN 1-58113-292-1.
13. PALMER, S., ROSCH, E., AND CHASE, P. Canonical perspective and the perception of objects. *Attention and Performance 9* (1981), 135–151.
14. PRESS, W., TEUKOLSKY, S., VETTERLING, W., AND FLANNERY, B. P. *Numerical Recipes in C*, 2nd ed. Cambridge Univ. Press, 1993.
15. RAMACHANDRAN, V., AND HIRSTEIN, W. The science of art a neurological theory of esthetic experience. *Journal of Consciousness Studies 6*, 6-7 (1999), 15–51.
16. SANDER, F. Gestaltpsychologie und kunsttheorie. ein beitrag zur psychologie der architektur. *Neue Psychologische Studien 8* (1931), 311–333.
17. SELIGMANN, D. D., AND FEINER, S. Automated generation of intent-based 3d illustrations. In *Proceedings of SIGGRAPH* (1991), pp. 123–132.
18. SOLSO, R. L. *Cognition and the Visual Arts*. MIT Press/Bradford Books Series in Cognitive Psychology, 1999.
19. VERFAILLIE, K., AND BOUTSEN, L. A corpus of 714 full-color images of depth-rotated objects. *Perception and Psychophysics 57*, 7 (1995), 925–961.
20. WEI HE, L., COHEN, M. F., AND SALESIN, D. H. The virtual cinematographer: A paradigm for automatic real-time camera control and directing. In *Proceedings of SIGGRAPH* (1996), pp. 217–224.
21. ZAKIA, R. D. *Perception and Imaging*. Focal Press Publications, 1997.

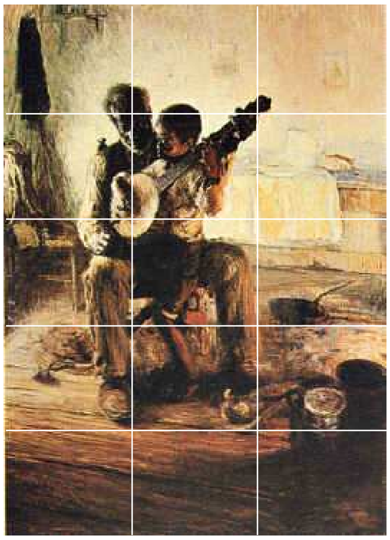


Fig. 3. The rules of thirds and fifths are examples of heuristic compositional rules. Linear elements often run along these lines and key features often occur at line intersections. (*Banjo Lesson*, Henry Tanner, oil on canvas.)

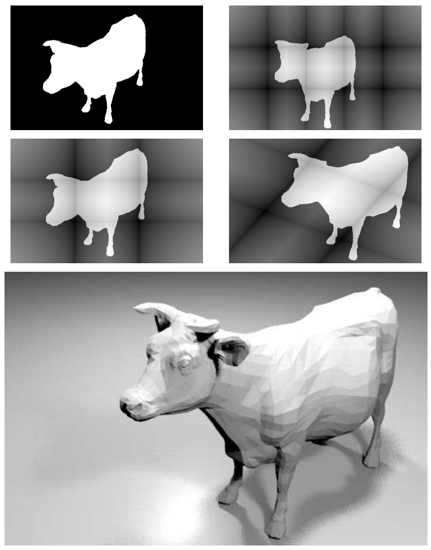


Fig. 4. Top left: initial viewpoint. Top right: combined rules of fifths and thirds. Middle left: rule of thirds. Middle right: angled rule of thirds. Bottom: rendered cow from angled rule of thirds.

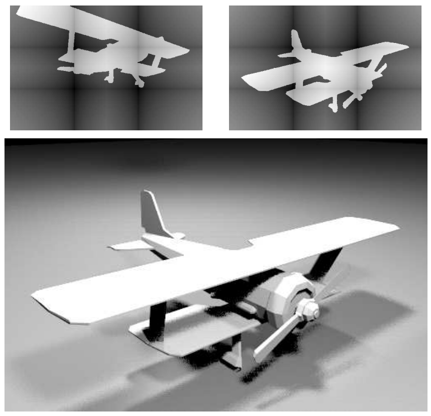


Fig. 5. Top: toy plane with rule of thirds layout and views from below and above. Bottom: toy plane rendered with view from above.

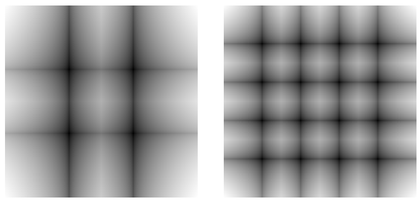


Fig. 6. Two images that guide layout optimization. The dark areas attract silhouette edges. The edges will tend to fall "downhill" toward these dark regions.

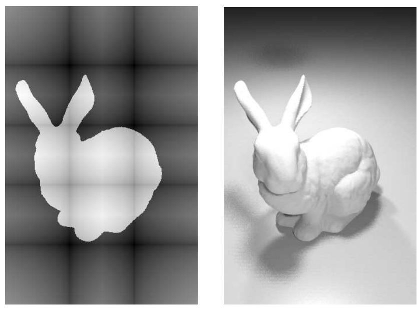


Fig. 7. Left: Bunny overlaid on a portrait format, combined rule of thirds and fifths template. Right: the resulting shaded image.