Automatic Image Creation via Artistic Composition Principles

Bruce Gooch Chris Moulding Donald Nelson Richard Riesenfeld Peter Shirley Brian Smits

University of Utah www.cs.utah.edu

Category: research

Format: print

Contact: Bruce Gooch
Department of Computer Science
University of Utah
50 S Central Campus Dr RM 3190
Salt Lake City, UT 84112–9205
phone: (801) 585–6161
fax: (801) 581–5843
email: bgooch@cs.utah.edu

Estimated # of pages: 8

Keywords: 3D Graphics, artistic rendering

Methods for choosing image parameters in both art and computer graphics are currently subjective. The choice of parameters results in images of varying quality. One aspect of image quality is the *composition* of the image. While the principles underlying composition are somewhat subjective, a portion of the compositional rules can be approximated quantitatively. We use this quantification to design an objective function and use numerical optimization to automatically arrive at images with acceptable composition. For a given subject or scene, the optimization procedure chooses format, viewpoint, layout, and lighting parameters. The resulting image is determined by characteristics the objective function rewards. We show several images generated using such optimization, and argue that these images have good composition.

Automatic Image Creation via Artistic Composition Principles

Category: research

Abstract

Methods for choosing image parameters in both art and computer graphics are currently subjective. The choice of parameters results in images of varying quality. One aspect of image quality is the *composition* of the image. While the principles underlying composition are somewhat subjective, a portion of the compositional rules can be approximated quantitatively. We use this quantification to design an objective function and use numerical optimization to automatically arrive at images with acceptable composition. For a given subject or scene, the optimization procedure chooses format, viewpoint, layout, and lighting parameters. The resulting image is determined by characteristics the objective function rewards. We show several images generated using such optimization, and argue that these images have good composition.

CR Categories: I.3.7 [Computing Methodologies]: Computer Graphics—3D Graphics

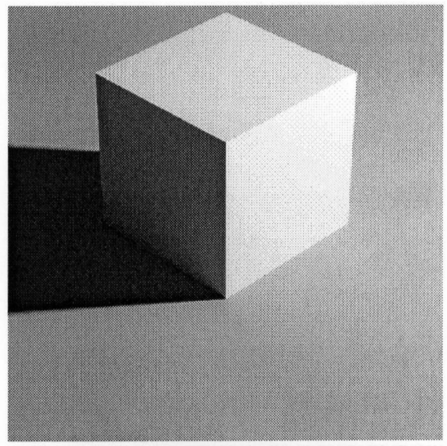
Keywords: 3D Graphics, artistic rendering

1 Introduction

This work is motivated by the extensive effort required to realize good composition in a computer generated image. Composition principles are simple and known to every artist, but their application is tedious and time consuming in many graphics situations. When an artist creates an image of a 3D object, the result is thought to have good or bad composition depending on how well the artist manipulates the free parameters of viewpoint, position, and framing. The choice of lighting is sometimes included under the umbrella of composition, and we adopt this broad definition. An example of two images of the same 3D object but with different compositional quality are shown in Figure 1. Several of the compositional principles that differentiate these images are measurable and can be accounted for algorithmically.

Little work dealing with artistic composition has been published in the computer graphics literature. Several researchers have attempted to optimize how well an image communicates [12, 16], others have borrowed principles from technical illustration [9, 21]. Some have examined how cinematographers develop animation sequences [11, 14]. Creating an environment with pleasing lighting has also been automated [15]. However, to our knowledge, there has been no work that directly deals with the question of how to organize a pleasing image automatically.

In this paper we attempt to automate some of the principles of composition to create "good" images of 3D objects. We draw compositional rules from both the art and psychology literature. It could be more correct to extract operational information entirely from psychology. However, although the psychology community has learned a great deal about some aspects of artistic composition [19, 22, 26], much of their knowledge is not yet specific enough to allow automation. Our method chooses view-angle, framing, and lighting as independent steps by incorporating artistic heuristics into an optimization process. We start with a 3D model of an object which is annotated with an up direction and a volume minimizing bounding box. We first decide on the size, shape, and orientation of a viewing window based on the proportions of the bounding box. Our algorithm then positions the object inside of the viewing window based on heuristics designed by artists to yield "interesting" divisions of the image plane, while maintaining an off axis view of the model. Finally we light the model using a lighting algorithm designed to increase the depth and shape information available in the image.



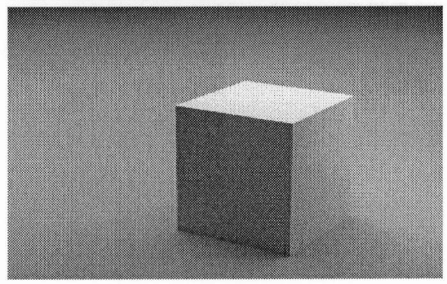


Figure 1: *Two images of the same object with different choices of free parameters. The bottom image is considered superior in its application of compositional rules because the frame is not square, the viewpoint is off-axis, the subject is placed off the horizontal and vertical centerlines, the shadows are subtle, and the lights have varying hues.*

2 Compositional Principles

In the art literature, the heuristics proposed by artists fall into four general categories for images of 3D scenes and objects:

1. choosing the format (image size, shape, and orientation),

2. choosing the viewpoint,

3. choosing the layout of the object or scene on the image plane,

4. choosing the lighting parameters.

Items 2 and 3 could both be considered choosing viewing parameters in computer graphics, but they are viewed as somewhat separate issues in the art community. In this section we discuss how artists optimize these four items in practice.

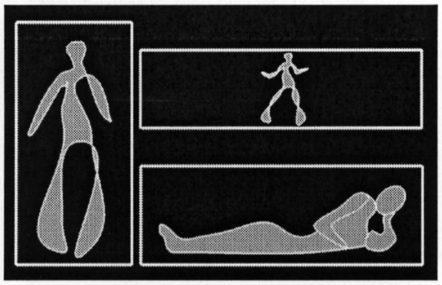


Figure 2: *The format of an image describes the shape and proportion of the image. The image on the left has a vertical format and the subject of the image is in accord with the format. Likewise, in the image on the bottom right, the horizontal figure is in harmony with its format. The vertical subject in the upper right image however, is out of relationship with the horizontal format and divides the image rather than becoming part of it.*

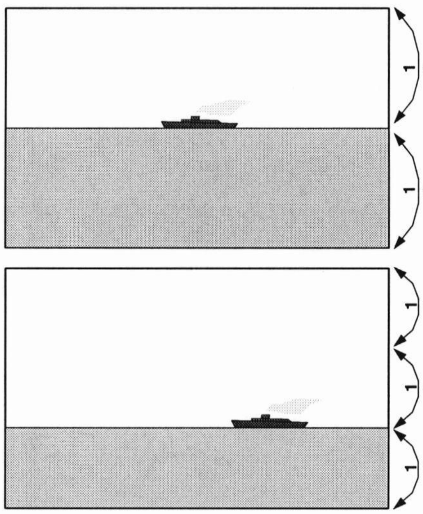


Figure 3: *Halving the canvas is considered inferior to dividing into thirds because halves create static compositions that can seem dull. Note that the rule is applied both horizontally and vertically (after Clifton [7]).*

2.1 Format

The format of a picture describes the size and shape of an image. An image that is wider than it is tall has a *landscape* format, otherwise it is has a *portrait* format. The format of a picture should be established at the beginning of a composition [5]. Landscape formats should be used with horizontal objects or scenes, and portrait formats with vertically composed images. This has the effect of allowing the scene to become part of the format rather than dividing it. Early work in psychology showed that the *golden ratios* seem to be preferred [20, 3]. The golden ratio is $(\sqrt{5}+1)/2 \approx 1.618$. A landscape image 500 pixels high would be about 809 pixels wide to follow this ratio. Artists often use a five by eight format, and this is usually regarded as owing to the similarity to the golden ratio.

2.2 Viewpoint

Psychologists have studied viewers' preferences for one viewpoint over another. A viewpoint that is preferred by most viewers is called a *canonical viewpoint*. Numerous experiments have investigated which viewpoints are likely to be canonical. Palmer *et al.* [17]

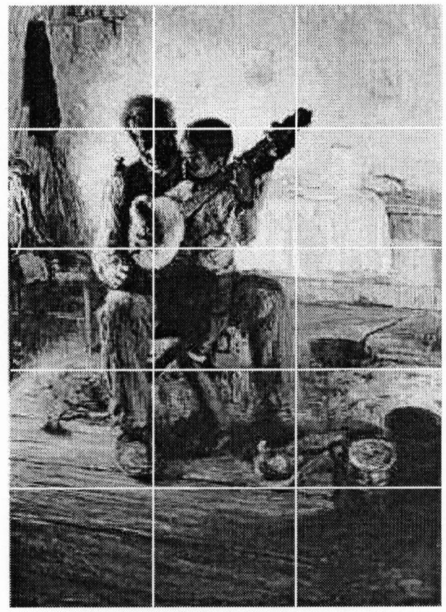


Figure 4: *The rules of thirds and fifths can be combined. Note that linear elements often run along lines and key features often occur at line intersections. A possible flaw in this composition is that the banjo strut might lead the viewer's eye to the corner. (*Banjo Lesson*, Henry Tanner, oil on canvas.)*

found that canonical views are off-axis. More recently, researchers discovered that a three-quarter view of a familiar object is preferred [24]. Canonical views for unfamiliar "nonsense" objects may also exist, but the data is not as strong [8]. Other studies have shown that humans prefer objects to be balanced in terms of their center of gravity [3].

A thorough investigation of canonical views was recently carried out by Blantz *et al.* [6]. They found three significant 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 human 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 are viewed from 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. When subjects in the study were given the ability to choose the viewpoint for an object Blantz *et al.* found that the subjects performed an internal optimization to find a viewpoint that showed the fewest number of occlusions. This occurred whether the objects were familiar or artificial geometric constructs. For instance, when choosing a viewpoint for a teapot the subjects always choose a viewpoint that showed both the handle and the spout.

2.3 Layout

Composition is taught to aspiring artists by showing them a few simple rules, then showing them a number of pitfalls to avoid. However, all sources agree that good composition can occur by bending the rules to the breaking point, or by just skirting the edge of the pitfalls.

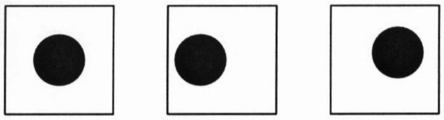


Figure 5: *The leftmost object is exactly centered yielding a solid but static quality. The middle object is considered too off center by standard artistic convention. The rightmost object's position is a compromise between too much symmetry and throwing the image off-balance.*

Rules of thirds and fifths. The best known rule of image layout is the *rule of thirds* (Figure 3). By partitioning their canvas into thirds both vertically and horizontally, and placing the strong vertical and horizontal portions of the image near these partitioning lines artists avoid equal spatial divisions of their image. Such equal spatial divisions will 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]. The rules of thirds and fifths can be mixed by dividing the canvas into thirds along one axis and fifths along the other (Figure 4).

Placement. Studies show that objects in a scene should be repelled from the corners and center of the format [2]. However, art theorists admit that the center of an image is where the most important information in the image should be placed [3, 22]. This seeming contradiction is explored in Figure 5. Subjects centered in the format become boring, but elements far off center are distracting.

Harmony of viewpoint and placement. Having chosen a viewpoint, it is good practice to place the subject in the bottom portion of the image if the viewpoint is above the subject with respect to gravity. Similarly it is a good idea to place the subject in the top portion of the image if the viewpoint is below the subject [7]. This is one of the differences in the images in Figure 1.

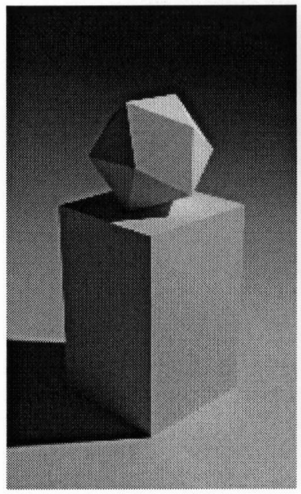
Feature orientation. Diagonal lines in an image yield a dynamic image. Strong horizontal and vertical lines tend to yield static images [7]. Lines oriented toward corners tend to draw the eye toward the corner and off of the image.

2.4 Lighting

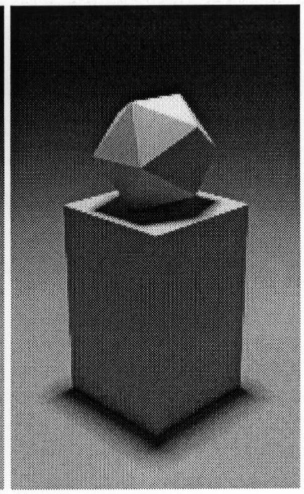
Lighting is used to give objects 3D shape definition, control the viewers attention, and sometimes to give a pleasing appearance. When dealing with complex scenes, lighting is a difficult task [1]. A lighting algorithim for a simple scene or a single object is more straightforward.

In photography light coming from the direction of the camera is known as front lighting because it illuminates the front of the subject. This method is used by fashion photographers since it tends to smooth out the subject, though this effect makes front lighting a poor choice for emphasizing shape. Likewise back-lighting, lights placed behind the subject pointing toward the camera, fails to give much shape information due to the visible sides of the subject falling into the shadow region. For these reasons photographers often use a combination of top and side lighting known as three quarter lighting.

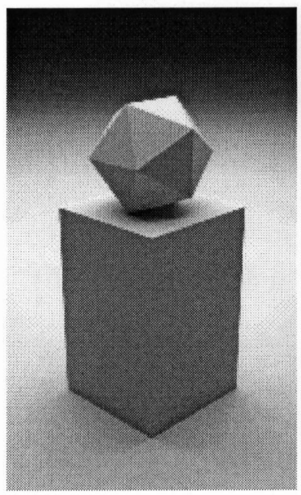
Placing the light source slightly behind the subject will result in a shadow cast slightly in front of the subject. This shadow will give the necessary depth and shading cues while at the same time adding a base for the subject to sit on. This type of shadow has little compositional impact. It also serves to darken the bottom of the subject in the image adding visual weight to the base of the subject. The main drawback to this lighting technique is that the shading across the subject tends to be dark instead of running a



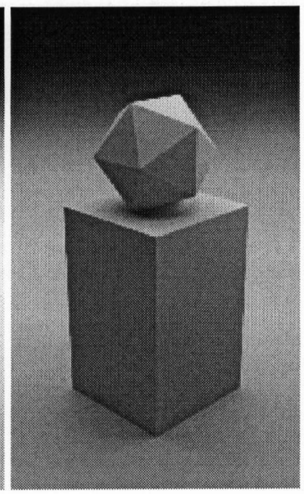
(a)



(b)



(c)



(d)

Figure 6: *Image (a) shows lighting from the side. This gives good definition, but the cast shadow is very prominent. Image (b) uses a light placed above and behind the subject, giving the benefit of a shadow at the base, but the shadow is distracting and hides detail. Image (c) increases the size of the overhead light, reducing the distracting quality of the shadow, but not solving the lack of detail problem. Image (d) includes a cool colored side light one quarter the size of the main light and changes the overhead light to a warmer color. This solves the detail problem.*

full intensity gamut. Photographers overcome this by placing a fill card, essentially a large white reflecting card, between the camera and the subject. Using colored lights is an additional method of adding shape information to an image [10, 18].

Another technique used by photographers to separate the subject from the background in their images is known as a falloff. Falloff is the result of uneven illumination in the background, and can be easily represented in computer graphics by a color blend from dark to light, or from a cool to warm color [13]. A technique similar to falloff for computer graphics was explored by Tanaka and Ohnishi [23].

Studio photographers face the practical problem of illuminating objects in a way that reveals shape without making the lighting visually dominate the image. Hunter and Fuqua [13] give an algorithmic way to light an object that avoids this problem. They derive their method incrementally as shown in Figure 6. Lighting with a small source to the side of an object reveals shape, but makes the shadow an important compositional element that draws the attention away from the subject. This is a common problem with shadows that should be avoided when the subject rather than the lighting is of primary importance [22].

3 Automatic Composition

To apply compositional principles in a computer graphics system, we use an optimization framework. We begin with a 3D model we would like an image of, and an initial set of image-generation parameters. The model does not have associated material parameters and we consider it to be a grey diffuse material. We have the following degrees of freedom: the position of the camera, the lighting, and the position of the subject when it is projected onto the image plane. Our implementation proceeds in two separate stages. The first chooses format, layout, and view parameters. The second sets up lighting parameters. We assume we know the direction of gravity relative to the model.

We use a golden rectangle for the image shape. We chose between portrait and landscape based upon whether the long axis of the bounding box is aligned with gravity. View and format parameters are specified as a 5D vector: (3D viewpoint, pan, tilt). Field-of-view is a fixed parameter usually set to that of a standard 35mm camera lens. We use an optimization procedure implemented in Matlab [25] to choose these parameters. The procedure encodes an objective function to evaluate the quality of a given composition, using an initial guess for our 5D vector.

The optimization attempts to find a good camera position based on the locations of the silhouettes on the screen. As the camera moves, the silhouettes change, so the first step inside the objective function is to take the edge list and find the silhouettes. The edges store both the location of the end points, and the normals of the two faces that share the edge. Once the silhouette edges are found, their midpoints are projected onto a $[0,1] \times [0,1]$ viewport. Points outside the viewport are not clipped, and contribute to the optimization. Four cost functions, C_c, C_f, C_tf, C_pt, of the n silhouette midpoints, (x_i, y_i), are weighted together to produce the cost of the current view. The first of these cost functions tries to keep the midpoints on the screen and is expressed as

$$C_c = \frac{1}{n}\sum_{i=1,n}(D * x_i - 1)^8 + \frac{1}{n}\sum_{i=1,n}(2 * y_i - 1)^8$$

where $D > 2$ pushes the object into the lower part of the image. Our value for D is 2.2. The second expression attempts to make the bounding box of the projected object fill a significant fraction A of the image area. This expression is as follows

$$C_f = ((\max_i x_i - \min_i x_i)(\max_i y_i - \min_i y_i) - A)^4.$$

We set $A = .6$ for our images. The third term tries to satisfy the rule of thirds or fifths (user controlled in each direction) by

$$C_{tf} = \sum_{i=1,n}\sqrt{.2(\min_{j=1,H} x_i - v_j)} + \sum_{i=1,n}\sqrt{.2(\min_{j=1,V} y_i - h_j)}$$

where H and V are either 3 or 5, and v_j and h_j are the positions of the lines. The initial viewpoint is chosen to fit the results of Blantz *et al.* [6]. We found the simple heuristic of choosing a viewpoint that constrains the area of the objects bounding box when projected to the screen to be (15% top, 55% largest side, 30% next largest side) produced seed viewpoints within the preferred range found by Blantz *et al.* [6] We think that any similar heuristic would work well. The initial view is assumed to be a reasonable view direction, and the goal of the optimization should be to fine tune the view direction while scaling and positioning the object on the canvas. For this reason the fourth expression of the objective function rewards solutions where the pan p and tilt t are close to the initial guess (p_i, t_i) as follows

$$C_{pt} = \sqrt{(p - p_i)^2 + (t - t_i)^2}. \quad (1)$$

The final objective function to be minimized is a linear combination of these terms as follows

$$C = 10C_c + 1000C_f + .02C_{tf} + C_{pt}. \quad (2)$$

Note that the weights in part result from the terms not being normalized, rather than being a direct measure of their priority.

Lighting is controlled using the algorithm illustrated in Figure 6. We place a large warm fill light above the object and a cool light beside the object. The cool light is one quarter the size of the large light.

Figures 7 and 8 show two runs of our algorithm, as well as default side views with diagonal lighting. Figure 9 shows a model of dolphins that has had some user-intervention after optimization. The optimized dolphins initially faced left. We made them face right, via a simple image flip, as is favored in cultures with left-to-right text [2]. We also added a cool ramp in the background to accent the warmly lit dolphins. Although these manipulations are simplistic, they illustrate the the type of user-intervention that could be made more sophisticated in the future. It is unlikely all such manipulations could be made automatic because semantic information is needed even for simple tasks such as determining front versus back. Also, the composition that would have the dolphins facing the viewer might be more desirable to some users, but again semantic information is needed for such manipulations.

4 Discussion

We have presented an overview of compositional principles and a proof-of-concept implementation that automates parameters for simple images based on some of the most quantitative compositional practices. There are many ways our method could be improved, such as changing the objective function in our optimization step, or by making our lighting algorithm more general. Also, we could optimize material properties and background parameters. We could also expand our objective function to penalize additional

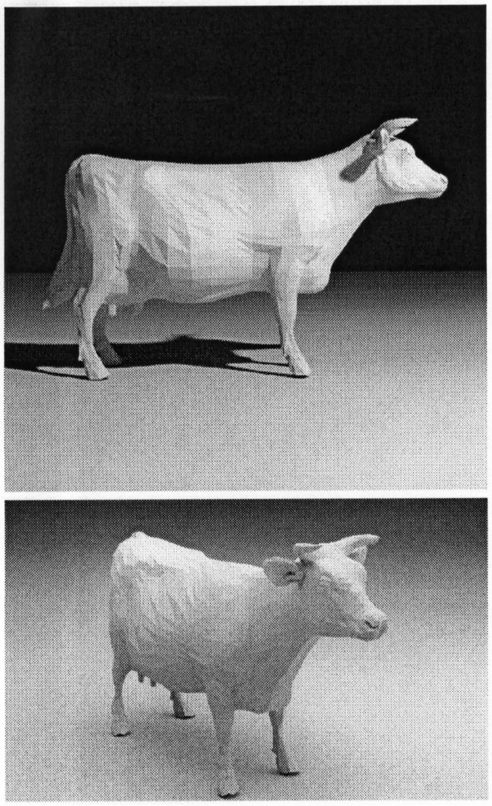


Figure 7: *Top: default composition. Bottom: optimized composition.*

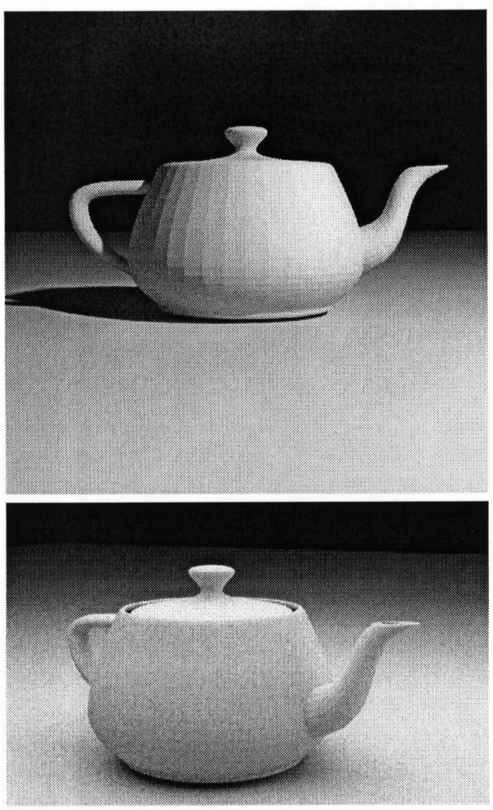


Figure 8: *Top: default composition. Bottom: optimized composition.*

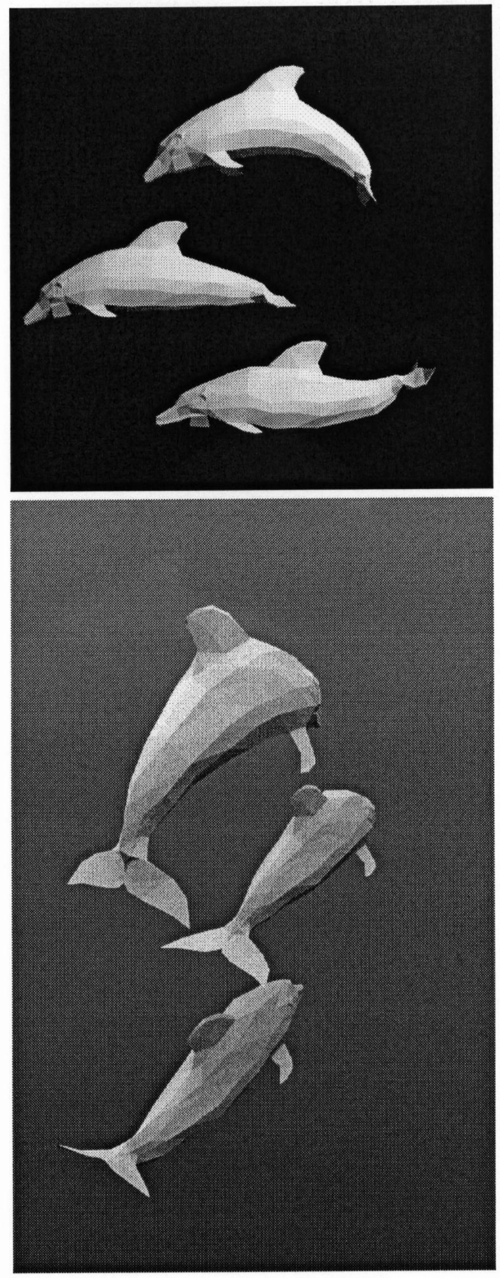


Figure 9: *Top: default composition. Bottom: optimized composition.*

"bad" compositions, for example, images that contain strong linear features pointed toward a corner can cause the viewers gaze to move off the image. Our current objective function operates on silhouette edges. These do not necessarily correspond to the most important image features. Better features could be extracted either from the object, or from a rendered image of the object. The penalty for these approaches, however, is that they must be done for each iteration of the optimization process.

Some aspects of composition are highly subjective and cannot easily be quantified. Grouping of objects in the image should be done in a manner which tells a story about the objects or describes their relationship with one another. Because this is highly dependent on the viewer and the subject it would be difficult if not impossible to automate. However, there are compositional rules that can serve as guidelines during the artistic process.

Gestalt psychologists have discovered methods which humans use to group objects in a scene. A survey of some of these high-level rules is given by Callahan [1]. There are also additional rules for placing multiple objects in a scene in order to lessen the chance of confusion on the part of a viewer [5, 19, 4]. These rules could be implemented as a type of "grammar check" for images. The most promising long-term approach is probably designing user-assisted systems that partially automate the image creation process.

This paper hangs on some good fortune. While advanced composition will doubtless remain the domain of the actively involved artist, quite acceptable and pleasing composition can actually be achieved, as we have shown, through applying a collection of basic principles representing the essential areas of concern for composition. We find the results both surprising and useful in a large category of situations where "good composition" is sufficient.

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 determmine canonical veiws. *Perception 28*, 5 (1999), 575–600.

[7] CLIFTON, J. *The Eye of the Artist*. North Light Publishers / Westport Conn., 1973.

[8] EDELMAN, S., AND BULTHOFF, H. Oreientation 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 and Applications 5*, 11 (Nov. 1985), 29–37.

[10] GOOCH, A., GOOCH, B., SHIRLEY, P., AND COHEN, E. A non-photorealistic lighting model for automatic technical illustration. *Proceedings of SIGGRAPH 98* (July 1998), 447–452. ISBN 0-89791-999-8. Held in Orlando, Florida.

[11] HE, L., COHEN, M. F., AND SALESIN, D. H. The virtual cinematographer: A paradigm for automatic real-time camera control and directing. In *SIGGRAPH 96 Conference Proceedings* (Aug. 1996), H. Rushmeier, Ed., Annual Conference Series, ACM SIGGRAPH, Addison Wesley, pp. 217–224. held in New Orleans, Louisiana, 04-09 August 1996.

[12] HORTON, W. Top ten blunders by visual designers. *Computer Graphics 29*, 4 (Nov. 1995), 20–24.

[13] HUNTER, F., AND FUQUA, P. *Light Science and Magic: An Introduction to Photographic Lighting*. Focal Press, 1997.

[14] KARP, P., AND FEINER, S. Issues in the automated generation of animated presentations. In *Proceedings of Graphics Interface '90* (May 1990), pp. 39–48.

[15] KAWAI, J. K., PAINTER, J. S., AND COHEN, M. F. Radioptimization - goal based rendering. In *Computer Graphics Proceedings, Annual Conference Series, 1993* (1993), pp. 147–154.

[16] LESTER, P. M. Digital literacy: Visual communication and computer images. *Computer Graphics 29*, 4 (Nov. 1995), 25–27.

[17] PALMER, S., ROSCH, E., AND CHASE, P. Canonical perspective and the perception of objects. *Attention and Performance 9* (1981), 135–151.

[18] PARRAMON, J. M. *The Book of Color*. Watson-Guptil Publications, 1993.

[19] RAMACHANDRAN, V., AND HIRSTEIN, W. The science of art a neurological theory of aesthetic experience. *Journal of Consciousness Studies 6*, 6-7 (1999), 15–51.

[20] SANDER, F. Gestaltpsychologie und kunsttheorie. ein beitrag zur psychologie der architektur. *Neue Psychologische Studien 8* (1931), 311–333.

[21] SELIGMANN, D. D., AND FEINER, S. Automated generation of intent-based 3D illustrations. In *Computer Graphics (SIGGRAPH '91 Proceedings)* (July 1991), T. W. Sederberg, Ed., vol. 25, pp. 123–132.

[22] SOLSO, R. L. *Cognition and the Visual Arts*. MIT Press/Bradford Books Series in Cognitive Psychology, 1999.

[23] TANAKA, T., AND OHNISHI, N. Painting-like image emphasis based on human vision systems. *Computer Graphics Forum 16*, 3 (August 1997), 253–260. ISSN 1067-7055.

[24] VERFAILLIE, K., AND BOUTSEN, L. A corpus of 714 full-color images of depth-rotated objects. *Perception and Psychophysics 57*, 7 (1995), 925–961.

[25] WOLFRAM, S. *The Mathematica Book*, 4th ed. Cambridge Univ. Press, 1999.

[26] ZAKIA, R. D. *Perception and Imaging*. Focal Press Publications, 1997.