

Creature Double Feature: On Style and Subject in the Art of Caricature

Kevin Burns

The MITRE Corporation
202 Burlington Road
Bedford, MA 01730-1420 USA
kburns@mitre.org

Abstract

A caricature is a distinctive drawing in which the essence of a subject is rendered in the style of an artist. The question is: How do artists make distinctive drawings and how do viewers see subject and style? Here I explore this question in a series of three investigations on contour drawings of animal silhouettes. First, I present a descriptive study of how different artists (styles) create freehand drawings of different animals (subjects). Next, I present a predictive model of how viewers compare contours when they recognize similarity in subject and style. Last, I present a generative program that creates caricatures using automated algorithms to simulate the cognitive processes of human artists. The resulting drawings are shown to exhibit and exaggerate the style and subject of human drawings. In closing, I discuss practical applications of caricature for improving human-system interaction and I outline possible extensions of contour analysis to other media like music and language.

Introduction

How do viewers recognize 3-D objects depicted in 2-D drawings (Burns 2001)? How do artists generate these drawings? What are the roles of *subject* and *style* in the viewer's process of perceiving objects and in the artist's process of creating drawings? These three questions are at the crux of how art works in drawing, and similar questions apply more generally to painting and sculpture as well as to music and language.

Here I focus on contour drawing as employed in the art of caricature, which some claim is the core of "all art" (Ramachandran and Hirstein 1999). In particular, the contour drawings that I study are animal caricatures, which are similar to the first forms of "wall art" (Mithen 1996; Wachtel 1993).

To set the stage, I outline a theory of vision, a theory of drawing and a theory of caricature. Based on these theories, I then present three studies on contour caricatures to address the questions noted above. In closing, I discuss how computer caricatures may apply to practical problems and how contour-based analysis may extend to non-visual media.

Theory of Vision

Here I adopt a *computational* theory of vision proposed by David Marr (1982), which postulates specific *representations* and *algorithms* in order to explain how people see things. The advantage of a computational theory is that it can make testable predictions in the form of computer simulations, which can then be compared to data collected from human experiments. The resulting comparison can shed light on both visual competence (i.e., what people can and cannot see) and visual illusions (i.e., when people are and are not fooled).

In Marr's theory, vision is a process that produces a series of internal images at increasing levels of abstraction. The input image is a pixel array (retinal projection), which is processed in early vision to produce a Primal Sketch. The Primal Sketch, hereafter called a *2-D Sketch*, is an internal image that represents major features of the input image as 2-D contours, much like a contour drawing but in the mind rather than on a page. The 2-D Sketch is then processed to produce a *2-1/2-D Sketch*, which is an internal image that represents higher-level features like the 3-D orientation of surfaces (whose boundaries project to 2-D contours in the Primal Sketch). Finally, the 2-1/2-D Sketch is processed to produce a *3-D Model*, which represents more abstract features of objects and their parts in an object-centered perspective (as opposed to a viewer-centered perspective like the 2-D and 2-1/2-D Sketches). Ultimately, recognition (naming) and association (recall) of objects is done with 3-D Models, which are the internal images that represent the most meaningful properties of the 3-D world.

Theory of Drawing

Assuming that an artist has an internal image of what he is trying to draw, the question is: What is the nature of this internal image (2-D, 2-1/2-D or 3-D) and how is it used to generate a drawing? The answer to this question obviously depends on the artist's goals and means. Here I focus on the case where an artist's goal is to make a *realistic* drawing and where his means include a *reference* image from which he draws. This can be contrasted to cases where an artist's goal is to make a schematic (abstract) drawing and/or where he does so from memory (recall).

When an artist looks at a reference image (or 3-D object), vision (per Marr's theory) generates a 2-D Sketch, 2-1/2-D Sketch and 3-D Model in a bottom-up (data driven) fashion. Of these three internal images, the 2-D Sketch is clearly the most useful one for realistic drawing since it explicitly represents the 2-D projections of 3-D contours, i.e., without extraneous information about surface orientation (in the 2-1/2-D Sketch) or object recognition (in the 3-D Model).

Although the 2-D Sketch is the most useful representation for realistic drawing (in art), it is the least useful representation for effective seeing (in life). This is because people need the 2-1/2-D Sketch to perceive 3-dimensional depth and surface orientation, and they need the 3-D Model to relate these 3-D dimensional features to objects. As such, realistic drawing can be seen as a rather unnatural activity, since to do so well one must ignore high-level representations that are normally very useful and instead focus on low-level representations that are normally not very useful (except insofar as they support the construction of higher level representations).

This theory explains why non-artists find it difficult to make realistic drawings, even when they are making drawings from a reference image (or 3-D object). That is, non-artists have simply not developed the skill of *seeing like an artist*, which is largely a matter of attending to the 2-D Sketch rather than the 3-D Model.

The theory also explains why artists prefer to draw from a reference image (or 3-D object) rather than from memory, at least when making realistic drawings. That is, in the absence of a reference image, the 2-D Sketch is relatively impoverished because it must be generated in a *top-down* fashion from the 3-D Model (via recall) rather than in a *bottom-up* fashion (via vision).

Finally, the theory explains why non-artists' drawings (esp. children's drawings) often appear *schematic* rather than *realistic* even when they are made from a reference image (or 3-D object). That is, the schematic features of these drawings represent more abstract features of the 3-D Model, which are hard to ignore because they are so useful in everyday activities.

In short, realistic drawing requires attention to the 2-D Sketch and suppression of the 3-D Model, which artists have learned to do and which non-artists (especially children) presumably find difficult to do (see Snyder and Thomas 1997; Humphrey 1999).

The above theory is in general agreement with a popular theory proposed by Betty Edwards (1979), based on the "split-brain" research of Roger Sperry (1973). Sperry's research shows that the right hemisphere is specialized for spatial tasks (like drawing) while the left hemisphere is specialized for language tasks (like naming). Edwards' theory holds that realistic drawing requires "turning off" the left (verbal) side of the brain, which is usually dominant, and "turning on" the right (visual) side of the brain. She proposes a number of specific exercises that people can use to do this, including the process of *contour drawing*, which has long been a standard approach in art schools (see Nicolaides 1941). In this process, an artist attends to the contour he is trying to draw (in a reference image) and slowly follows along it with his gaze (in the reference image) while at the same time drawing it on the page (in a freehand drawing). According to Edwards, this technique suppresses the left brain and activates the right brain and thereby helps people make better (more realistic) drawings.

The problem with Edwards' theory is that it is a motivational theory rather than a computational (or biological) theory. That is, she does not specify what kinds of representations or algorithms are used by each side of the brain, let alone how these representations and algorithms are implemented in neural hardware. As such, the right-brain theory serves mainly to provide new-age reasons (from split-brain research) for trying age-old techniques (like contour drawing). In contrast, a computational theory like Marr's can be used to build and test scientific models with more descriptive, predictive and generative power. And that is my purpose here.

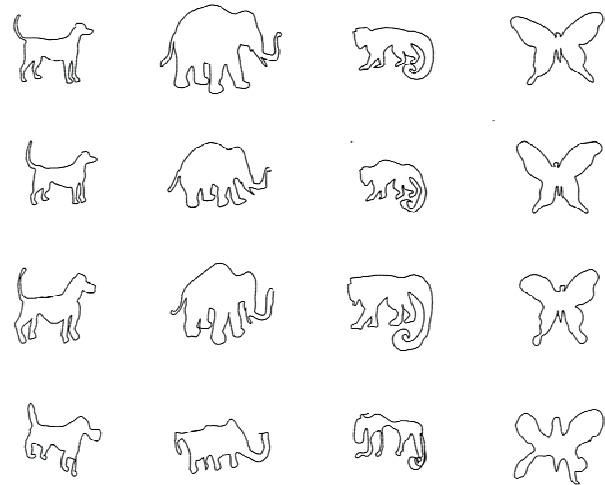


Figure 1. Animal cartoonettes were the stimuli (row 1) and responses (rows 2-4) in a descriptive study of freehand drawing. Drawings by different artists (styles) are in different rows. Drawings of different animals (subjects) are in different columns.

Theory of Caricature

A caricature is a drawing that exaggerates the *distinctive* features (Rhodes and McLean 1990) of a subject. For example, if a person has a crooked smile then a caricature will render it even more crooked than it really is. Much has been written about why caricatures are effective from an aesthetic perspective (Gombrich 1960; Ramachandran and Hirstein 1999) and about how caricatures are created from a practical perspective (Gautier 1985; Redman 1984). But not much has been written about the workings of caricature from a computational perspective, except for the research of Susan Brennan (1985; also see Rhodes, Brennan and Carey 1987; Bruce and Young 1998).

Brennan's research is novel in that she proposes a theory of caricature and implements a model of it in computer simulations. In Brennan's theory, the generation of a caricature requires three things: (1) a contour drawing of a subject face (to be caricatured), (2) a contour drawing of a reference (average) face, and (3) the point-wise correspondences between contours of the subject face and the reference face. In Brennan's model, the caricature is then generated by: (1) defining the vector difference between each pair of corresponding points, (2) increasing the distance along this vector to create a new point, and (3) connecting these new points to obtain the contours of the caricature.

Brennan's model is remarkably effective in exaggerating the distinctive features of human faces, as judged by human viewers. However, the model is not a complete test of the theory because the resulting caricatures are not completely the work of the computer. That is, Brennan establishes point-wise correspondences herself (by hand), which begs the question: How much of the art lies in the hand of the human and how much has been captured by the automated algorithm? My investigation (below) is novel in that I develop and implement automated algorithms for all steps of computerized caricatures, albeit for simpler stimuli than human faces.

Investigations

A Descriptive Study: Drawing Distinctions

My first investigation was a descriptive study of *realistic* drawings made from a *reference* image (2-D). This study differs from previous research, which has focused on adult drawings of schematic figures or children's drawings made from either 3-D objects or memory (see van Sommers 1984; Reith 1988; Willats 1997; Lange-Küttner, Kerzmann and Heckhausen 2002). I studied three artists, two adults and an 8-year old child, using contour drawings of animal silhouettes that I call *cartoonettes* (Figure 1).

Each artist was presented with a reference image (top row of Figure 1) and asked to draw the best freehand copy that he/she could. All three artists (Figure 1 rows 2-4) made drawings of all four animals (Figure 1 columns 1-4).

The reference images (row 1 of Figure 1) were shown on a computer screen, one at a time, to each artist. The artist made his or her freehand drawing on a digitizing tablet while the in-progress drawing was displayed beside the reference image on the computer screen. Artists were told to use the technique of *contour drawing* (see above), where they slowly follow (with their gaze) along the contour of the reference image while making their drawing on the digitizing tablet (with their pen).

The artists were allowed to shift their gaze from the reference image to their freehand drawing as often as they wished in order to compare the contours, and to pause their pen as often as they wished in order to reflect on their progress. The only requirement was that they keep the pen on the page (tablet) at all times to ensure that the final drawing was a single contour.

This experiment was intended to explore how well people (adults and children) perform in contour drawing, and in particular to gain insight into how well they can attend (in their mind) and render (on the page) the internal image referred to above as the 2-D Sketch. The experiment was designed to test these abilities as directly as possible, by casting both the stimulus and response in the form of a 2-D Sketch – i.e., contour drawing.

In the more complex case where an artist draws from a 3-D object, the drawing process requires three steps: (1) to *see* the 3-D object as a 2-D image (projection), akin to the pixel array (input image) in Marr's theory, (2) to *find* the 2-D contours that will be represented in the final drawing, akin to the 2-D Sketch in Marr's theory, and (3) to *draw* the 2-D contours that are represented in the 2-D Sketch, using a pen on a page. Previous research on children's drawings made from 3-D objects (van Sommers 1984; Reith 1988; Willats, 1997), involving all three steps, shows that much of the rendered contour is schematic rather than realistic (for children under the age of 10).

My results in row 4 of Figure 1, although for only one child, show a remarkable amount of realistic contour (i.e., accurate copying) rather than the schematic *circle and stick* contours usually exhibited by children. My task requires only step 3 (see above), so the result suggests that the major hurdles to realistic drawing in children (and adults) are at steps 1 and 2 where one must *see* 3-D things in 2-D terms and *find* the lines to draw. In fact, even expert artists often use tricks to ease these steps, by drawing from photographs (to bypass step 1) or by copying images created by other artists (to bypass steps 1 and 2).

The artists in Figure 1 only had to *draw* the lines (step 3), but the results still exhibit some interesting features of subject and style. With respect to subject, each animal is rather easily identified in the adults' drawings (row 2 and row 3), as it is in the reference images (row 1). As noted above, even the child's drawings (row 4) exhibit a remarkable amount of realistic contour and hence are somewhat recognizable. Thus, by isolating step 3 of the drawing process, we see that both adults and children are capable of suppressing the 3-D Model and accessing the 2-D Sketch in contour drawing, but to different degrees.

With respect to style, the child's drawings (row 4) clearly exhibit more aspects of schematic contour than the adults' drawings (rows 2 and 3). For example, the legs in row 4 are *stick-like* in that they are more uniform in length, width, shape (parallel) and space (between legs) than the realistic leg contours in the reference image. This suggests that the child has more trouble suppressing her high level representation (3-D Model) and attending to a low level representation (2-D Sketch), although motor skills (independent of attention) may also play a role.

In fact, one might use the amount of realistic versus schematic contour to characterize the *style* difference between adults and children. But to do so one would need formal measures of *realistic* versus *schematic* contour (see Reith 1988; Lange-Küttner, Kerzmann and Heckhausen 2002) in order to develop computational metrics for qualitative labels such as *stick-like*. A first step in this direction, using a simple method based on contour matching, is discussed below under "A Predictive Model".

Also with respect to style, Figure 1 shows that there are noticeable differences between the realistic contours produced by the two adults. For example, the artist in row 2 tends to exaggerate the *sharpness* of the legs, snouts and tails, making some rather pointed contours compared to the reference contours. In contrast, the artist in row 3 tends to exaggerate the *thickness* of the legs, snouts and tails, making them look rather smooth and fat. These features appear across all animals in a row, hence they can (like the child's tendency to produce schematic contours) be considered features of artistic style (not animal subject).

Finally, with respect to the relation between style and subject, the freehand drawings in Figure 1 suggest that there is no clear-cut distinction. This is because an artistic style that exaggerates the curvature of feet or the thickness of legs may also exaggerate distinctive features of the animal subject. For example, elephants have fat legs compared to dogs, and this is particularly prominent (caricatured?) in the elephant of row 3 even though all of the drawings in row 3 have relatively fat legs compared to their reference images in row 1. Thus, while it may be relatively easy to distinguish style (artist) from subject (animal) in the *forward* generation of drawings (i.e., in synthesis, see Freeman, Tenenbaum and Pasztor 2003), it may be much harder to distinguish the two in the *reverse* recognition of drawings (i.e., in analysis, see Tenenbaum and Freeman 2000).

A Predictive Model: Subject and Style

The descriptive study of freehand drawing (Figure 1) illustrates differences in both subject (between columns) and style (between rows). The question now is: How do viewers identify these differences in order to recognize the animal's subject and distinguish it from the artist's style? Here I present an objective (mathematical) model of the subjective (psychological) process, which involves piecewise comparison of a subject contour to a reference contour in order to judge the *similarity* between these two contours.

My matching model works by comparing the orientations of contour segments. The idea is to approximate the curved contour R of a reference image by straight segments $R_1, R_2, R_3, \dots, R_N$, and similarly to approximate the curved contour S of a subject image by straight segments $S_1, S_2, S_3, \dots, S_N$. Then, the following *constrained cost* algorithm is used to establish the correspondences between R and S:

Matching Model

- (1) Each curved contour (R and S), which typically contains about 1000 pixel points, is first smoothed by a Gaussian filter to reduce the noise. Each smoothed contour is then approximated by $N=200$ straight segments. All R_i are of equal length and all S_j are of equal length. The 2-D orientation (dx, dy) of each segment R_i and S_j is computed and stored in vectors $O(R_i)$ and $O(S_j)$.
- (2) The segments R_1 and S_1 are taken as the first match. The indices are incremented, setting $i=2$ and $j=2$.
- (3) The cost is computed for various choices that can be made for the next match. The choices are constrained to only the following three possibilities:
 - a. (i, j) : $\text{Cost}(R_i, S_j) = [O(R_i) - O(S_j)]^2$
 - b. $(i+1, j)$: $\text{Cost}(R_{i+1}, S_j) = [O(R_{i+1}) - O(S_j)]^2$
 - c. $(i, j+1)$: $\text{Cost}(R_i, S_{j+1}) = [O(R_i) - O(S_{j+1})]^2$
- (4) Of these three choices, the one with the least cost is chosen at the new match.
- (5) Given the new match, the indices are incremented and steps 3 and 4 are repeated until the end of contour R or contour S (or both) is reached.
- (6) The total cost of the match between R and S is computed as the sum of the match costs for the individual segments.

Figure 2 illustrates the resulting match for two drawings of the dog. One contour is the reference image (smoothed) and the other contour is an artist's drawing (smoothed). The connecting lines illustrate the computed correspondences. This figure shows that the simple model (above) does a good job of matching the corresponding points of high curvature (nose, feet, tail, etc.), which are known to play a key role in human vision (Richards, Dawson and Whittington 1988).

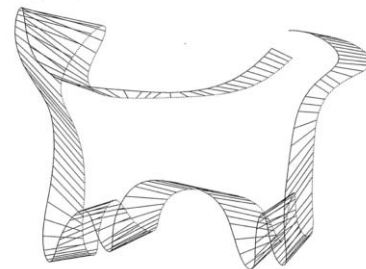


Figure 2. Ribbon plot shows results of matching model.



<u>00</u>	83	70	77	83	<u>00</u>	200	083	70	200	<u>00</u>	82	77	106	082	<u>00</u>
<u>01</u>	58	62	81	75	<u>10</u>	174	116	64	140	<u>02</u>	73	64	070	070	<u>01</u>
<u>04</u>	67	83	73	75	<u>07</u>	220	137	79	155	<u>05</u>	80	69	090	104	<u>02</u>
<u>09</u>	52	87	86	68	<u>35</u>	139	122	83	138	<u>12</u>	62	48	106	098	<u>12</u>

Figure 3. Cost scores computed by the matching model. Each number in a matrix is the cost of matching the corresponding (row, column) drawing in Figure 1 to a reference image (shown above the matrix). Bold-underlined numbers are the best matches (lowest cost).

Besides finding piecewise correspondences, the matching model also computes the overall goodness (cost) of the match between R and S. This provides a predictive measure of similarity between R and S, which can then be compared to human judgments in order to test the psychological plausibility of the model.

Figure 3 shows the computed cost of matching each drawing of Figure 1 to the four references images. The bold-underlined numbers show that the model can clearly discriminate the good matches (same animal) from the bad matches (different animal) – i.e., to recognize the *subject* of a drawing across columns.

Figure 3 also shows that the matching model makes specific predictions, between rows, about the similarity of different *styles* for a given subject. Of course, the match cost is only a single (global) measure that does not say much about the many (local) features that might be used to distinguish one style from another. For example, the qualitative features of the various leg styles noted earlier in Figure 1 (i.e., *sharpness* in row 2, *thickness* in row 3 and *stick-like* in row 4) might be considered intermediate features of style that lie between the very local difference in orientation (computed at each step of the matching process) and the very global feature of total cost (computed at the end of the matching process). In fact, I suspect that intermediate features play a key role in how people recognize different *subjects* as well as different *styles*. For example, people may weigh the match of certain contour segments (e.g., *head* or *legs* or *tail*) more than other segments in recognizing the identities of animals.

Obviously the measuring and modeling of such intermediate features requires some *top-down* processing (see Cavanagh 1991), e.g., the head must be recognized before one could more heavily weigh the contour segments that comprise the head. However, these top-down processes are governed by an object’s meaning, so they are much more dependent on knowledge of 3-D objects and much less dependent on features of 2-D images. In short, top-down processing is knowledge driven, while bottom-up processing is data driven. Here my approach is to avoid top-down modeling, both to keep the study tractable and to see how much can be done from the bottom up.

Thus, with a focus on *globally* measured features computed in a *bottom-up* fashion, I extended my investigation of style by having one artist make more drawings. This artist (in row 2 of Figure 1) was asked to make three more drawings of one reference image (the dog) in different styles. I did not specify what I meant by “different styles”, and the artist did not ask, but he then proceeded to make the three drawings shown in the middle row of Figure 4. These drawings indeed exhibit styles different from the previous drawing by this artist (shown as number 1 in Figure 4). The three new styles might be characterized, moving left to right in the middle row of Figure 4, as *husky*, *skinny* and *pointy*. Recall that the three styles in the bottom row, which were generated by three different artists, were characterized (moving left to right) as *stick-like*, *thickness* and *sharpness*.

Of course, these style labels are subjective and other people might assign different labels to highlight other features (global or local) of an artist’s style. Here I offer the labels only as a convenience for referring to different drawings, as I compare the results of my matching model more formally to the judgments of human viewers below.

Assuming that the cost computed by the matching model can be used as a global measure of style, I computed the match cost for all six of the dog drawings (against the reference image, which is show as the top dog in Figure 4). The dogs were then ranked one to six (best to worst) based on the cost (low to high) computed by the matching model. The computed ranks are noted in Figure 4.

For comparison I collected data from 27 human judges. The judges were shown Figure 4 (without the ranks of the matching model) and asked to “rank each dog drawing from one to six, based on its similarity to the top dog”. The measured ranks are compiled in Figure 5.

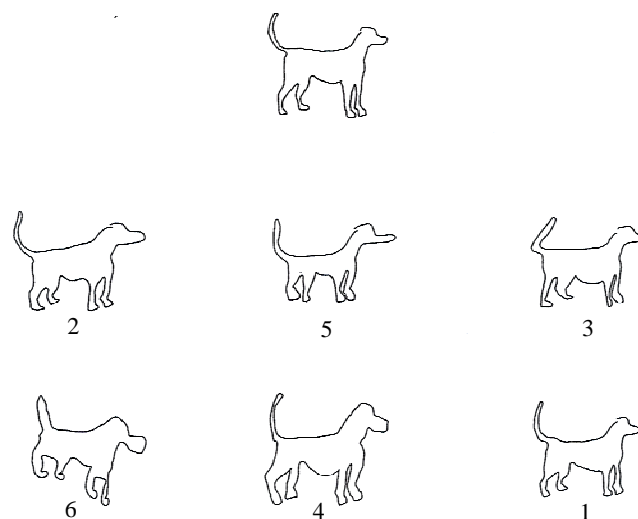
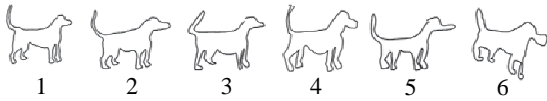


Figure 4. The dog drawings (numbered 1-6) are compared to a reference image (top dog). The numbers are rankings predicted by the matching model.



1	25	1	0	1	0	0
2	1	13	9	2	2	0
3	1	5	15	6	0	0
4	0	5	2	17	3	0
5	0	3	1	1	20	2
6	0	0	0	0	2	25

Figure 5. Ranking results from human judges. Each cell (row, column) shows the number of people whose assigned rank (in row) corresponds to the model's rank (in column).

The high numbers along the diagonal of Figure 5 show that the rankings by human judges are in accord with the matching model. The human judges were almost unanimous for the dogs in columns 1 and 6, and the human rankings were also very modal for the dogs in the other columns. For 11 of the 27 judges, the reported rank for all 6 dogs matched the predicted rank exactly. In a test of statistical significance using Kendall's coefficient of concordance (W), $W = 0.78$, $\chi^2(5) = 105.3$, $p < 0.001$.

These results for *style* (Figure 5), which are a stronger test of the matching model than the results for *subject* (Figure 3), suggest that the model provides a plausible simulation of how people compare contours when they judge *similarity*. The model also provides a useful method of separating subject from style, simply by using an empirical threshold for what constitutes a good match. For example, as seen by the numbers in Figure 3, a threshold of 40 can distinguish subject from style for all of the freehand drawings in Figure 1.

It is important to note that my study was limited to only a few different subjects (animals) and a few different styles (artists). And, while my global and bottom-up approach may extend to other subjects and styles, it is not clear how well the approach will generalize, for three reasons. First, establishing style differences like *sharpness*, *thickness* and *stick-like* (or *husky*, *skinny* and *pointy*) requires the computation of intermediate (between local and global) features that can be used to distinguish such styles. Second, the recognition of both style and subject involves top-down processing, i.e., where the meaning of a drawing affects the bottom-up vision. And third, there is no a priori reason to assume that distinctions between style (artist) and subject (animal), which may be separable in generating a drawing, will still be recoverable in recognizing a drawing.

In light of these limitations I turn now to my third investigation, which focuses on the process of generating (synthesizing) style and subject together – rather than recognizing (analyzing) style and subject differences. The question is: Can automated algorithms be used to generate caricatures that exhibit and exaggerate the style and subject of human drawings?

A Generative Program: Computer Caricatures

According to theories of caricature (above), the drawings in Figure 1 can be considered caricatures of animals (subjects) by artists (styles). Also according to theories of caricature, the workings of these drawings as caricatures in the minds of viewers and artists depend on *matching* (correspondences) in vision and *making* (exaggerations) in drawing (Gombrich 1960).

My predictive study (above) showed that a matching model could predict human judgments of similarity between two contour drawings. The idea now is to use the matching model as a central component of an automated algorithm for creating caricatures, and thereby overcome the major limitation of previous computer caricatures where point-wise correspondences were *handpicked*.

Here I assume that when an artist draws a caricature from a reference image he actually uses two internal images. One image is a 2-D Sketch of the reference image, hereafter called the *subject* image, which is generated in a bottom-up fashion. The other image is a 2-D Sketch of an *average* subject, which represents a prototype whose generation involves some top-down processing (i.e., since an average is a generalization that must be recalled from memory). The creation of a caricature then involves three steps: (1) *recalling* the average image (2-D Sketch, generated top down), (2) *comparing* the subject image (2-D Sketch, generated bottom up) to the average image, and (3) *creating* an output image, which exaggerates the distinctive differences between the subject and the average.

These three steps can be simulated in a purely bottom-up fashion, using my matching model (for step 2) along with Brennan's exaggeration equation (for step 3), if the average image (step 1) can be obtained without top-down processing. Thus, to make it all bottom-up, I simply assume that the *reference* image (top row of Figure 1) is a reasonable surrogate for the *average* image. [Alternatively, as done by Brennan and others, one can compute an average image from a large number of subject images.]

With the above assumptions, my model generates computer caricatures using the following automated algorithm:

Computing Caricatures

- (1) Two contour drawings are provided as input. One is a reference contour R and the other is a subject contour S (to be caricatured).
- (2) The matching model (see above) is used to establish piecewise correspondences between R and S . Each correspondence k defines a pair of points, $R_k(x, y)$ and $S_k(x, y)$ where x and y are the image coordinates of the corresponding points.
- (3) An exaggeration equation (per Brennan) is used to define new points C_k , which have the following x and y coordinates (where "e" is an exaggeration parameter):

- a. $C_k(x) = R_k(x) + e * [S_k(x) - R_k(x)]$
 b. $C_k(y) = R_k(y) + e * [S_k(y) - R_k(y)]$
- (4) The points $C_k(x, y)$ are connected (in the same order as they are in S and R , i.e., $k=1, 2, 3$, etc.). The resulting contour is the computed caricature.

Figure 6 shows the computed caricatures (C) for all six of the human dog drawings (S) relative to the reference dog image (R). Caricatures are shown for two different values of the exaggeration parameter, $e=1.25$ and $e=1.50$.

While high e can make the drawings look distorted, most viewers (in informal polls) say that the modest caricatures (e.g., with $e=1.25$) are *more distinctive* of animal subject and artist style than either the reference images or the human drawings. In more formal experiments on distinctiveness, Rhodes and McLean (1990) used Brennan's method of computing caricatures (by handpicking correspondences) to study contour drawings of bird types and found that caricatures were recognized more quickly than "veridical" line drawings. In their study, the optimal (preferred) degree of exaggeration was about 40%. Similarly, for contour drawings of human faces, Rhodes, Brennan and Carey (1987) found that caricatures were recognized faster than non-caricatures, which were recognized faster than anti-caricatures (negative e), with average times of 3 sec., 6 sec. and 12 sec., respectively (and with an optimal level of exaggeration of about 20%).

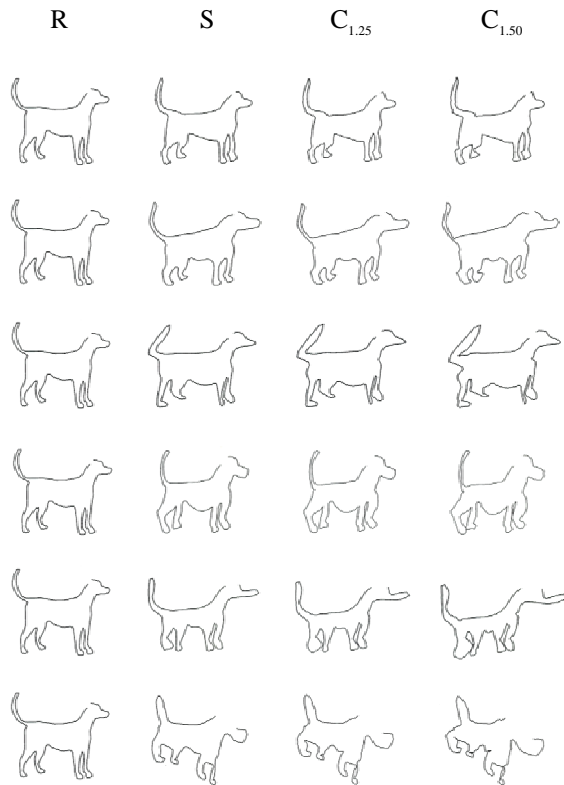


Figure 6. Computer caricatures (C) of the six human drawings (S) relative to the reference image (R).

Conclusion

This paper investigated the relation between style and subject in art, with a focus on contour and caricature. I would emphasize the following contributions: (1) Basic theories of vision, drawing and caricature, which can explain human competence in visual processing. (2) Working models of contour matching and caricature generation, which simulate human judgments of *similarity* and which exaggerate the *distinctiveness* of human drawings. I would also acknowledge the following limitations: (1) The computational models apply only to global and bottom-up aspects of contour and caricature. (2) The experimental results are limited to a small set of subjects (animals) and styles (artists).

The theories outlined in this paper are useful because they can guide the engineering of *sketching systems* and *data displays*. Computer systems and their visual displays are often designed without underlying theories of what users want or need to do their work – and better theories are needed to engineer systems and displays that can assist users in useful ways, e.g., by *amplifying the mind's eye* (Fish and Scrivener 1990).

The models proposed in this paper are useful because they were shown to be *psychologically plausibly* and because they were shown to draw caricatures using fully *automated algorithms*. Previous demonstrations of computer caricatures have relied on handpicking of point-wise correspondences, which is tedious and time consuming. A fully automated approach, like that outlined here, is needed if the advantages of computer caricatures are to be exploited in applications aimed at achieving *sketch beautification* or enhancing *sketch recognition* in human-computer interaction.

With further extensions to overcome current limitations (e.g., by considering multiple contours and by implementing some knowledge structures for top-down processing), computer caricatures might be generated for various applications in web pages and printed media. Contour-based caricatures (line drawings) are particularly useful in this regard because they have low bandwidth and because they are often (perhaps because of low bandwidth?) more readily recognized by human viewers than photographs or other more detailed imagery.

Finally, although I have focused on vision and drawing, a similar approach may apply to other modalities (see Kennedy 1993) and other media. For example, the *lines of melody* in music can be represented as pitch-time contours (Seashore 1967), and the musical tradition of *theme and variations* can be considered a sort of caricature. In fact, some visual artists have literally adopted this analogy in their work, as in the contour drawings of Matisse (1995). The major difference, of course, is that contour in music is temporal, while contour in drawing is spatial. Nevertheless, drawings (like music) are made in time.

Thus, computer-based studies of pitch-time contour in music may shed light on the process of drawing, which can be considered a sort of visual music. Conversely, computer-based studies of caricature in drawing may inform the study of pitch-time contour in music.

Similarly, the *lines* of poetry in language include stress-time contours. However, unlike drawing and music where contour is largely *signal-like* and hence can be analyzed in a bottom-up (data driven) manner, language is largely *symbolic* and so it must be analyzed in both a bottom-up and top-down (knowledge driven) manner. A contour-based approach to language would be most useful for recognizing and generating *style* (manner), as expressed in the signal-like contours of stress-time inflection, rather than *subject* (matter).

References

- Brennan, S. E. 1985. Caricature Generator: The Dynamic Exaggeration of Faces by Computer. *Leonardo* 18(3):170-178.
- Bruce, V., and Young, A. 1998. *In the Eye of the Beholder: The Science of Face Perception*. Oxford, UK: Oxford University Press, 165-176.
- Burns, K. 2001. Mental Models of Line Drawings. *Perception* 30:1249-1261.
- Cavanagh, P. 1991. What's Up in Top-Down Processing. In A. Gorea, ed. *Representations of Vision: Trends and Tacit Assumptions in Vision Research*. Cambridge, UK: Cambridge University Press, 295-304.
- Edwards, B. 1979. *Drawing on the Right Side of the Brain*. Boston: Houghton Mifflin.
- Fish, J., and Scrivener, S. 1990. Amplifying the Mind's Eye: Sketching and Visual Cognition. *Leonardo* 23(1):117-126.
- Freeman, W. T., Tenenbaum, J. B., and Pasztor, E. 2003. Learning Style Translation for the Lines of a Drawing. *ACM Transactions on Graphics* 22(1):33-46.
- Gautier, D. 1985. *The Art of Caricature*. New York: Perigee Books.
- Gombrich, E. H. 1960. *Art and Illusion: A Study in the Psychology of Pictorial Representation*. Princeton, NJ: Princeton University Press.
- Humphrey, N. 1999. Cave Art, Autism and the Evolution of the Human Mind. *Journal of Consciousness Studies* 6:116-143.
- Kennedy, J. M. 1993. *Drawing and the Blind: Pictures to Touch*. New Haven: Yale University Press.
- Lange-Küttner, C., Kerzmann, A., and Heckhausen, J. 2002. The Emergence of Visually Realistic Contour in the Drawing of the Human Figure. *British Journal of Developmental Psychology* 20:439-463.
- Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W. H. Freeman.
- Matisse, H. 1995. *Drawings: Themes and Variations*. New York: Dover.
- Mithen, S. 1996. *The Prehistory of the Mind: The Cognitive Origins of Art and Science*. New York: Thames and Hudson.
- Nicolaides, K. 1941. *The Natural Way to Draw*. Boston: Houghton Mifflin.
- Ramachandran, V. S., and Hirstein, W. 1999. The Science of Art: A Neurological Theory of Aesthetic Experience. *Journal of Consciousness Studies* 6:15-51.
- Redman, L. 1984. *How to Draw Caricatures*. Chicago: Contemporary Books.
- Reith, E. 1988. The Development of Use of Contour Lines in Children's Drawings of Figurative and Non-Figurative Three-Dimensional Models. *Archives de Psychologie* 56: 83-103.
- Rhodes, G., and McLean, I. A. 1990. Distinctiveness and Expertise Effects with Homogeneous Stimuli: Towards a Model of Configurational Coding. *Perception* 19:773-794.
- Rhodes, G., Brennan, S., and Carey, S. 1987. Identification and Ratings of Caricatures: Implications for Mental Representations of Faces. *Cognitive Psychology* 19:473-497.
- Richards, W., Dawson, B., and Whittington, D. 1988. Encoding Shape by Curvature Extrema. In W. Richards, ed. *Natural Computation*. Cambridge, MA: MIT Press, 83-98.
- Seashore, C. E. 1967. *Psychology of Music*. New York: Dover.
- Snyder, A. W., and Thomas, M. 1997. Autistic Artists Give Clues to Cognition. *Perception* 26:93-96.
- Sperry, R. W. 1973. Lateral Specialization of Cerebral Function in the Surgically Separated Hemispheres. In F. J. McGuigan and R. A. Schoonover, eds. *The Psychophysiology of Thinking*. New York: Academic Press, 209-229.
- Tenenbaum, J. B., and Freeman, W. T. 2000. Separating Style and Content with Bilinear Models. *Neural Computation* 12(6):1247-1283.
- van Sommers, P. 1984. *Drawing and Cognition: Descriptive and Experimental Studies of Graphic Production Processes*. Cambridge, UK: Cambridge University Press.
- Wachtel, E. 1993. The First Picture Show: Cinematic Aspects of Cave Art. *Leonardo* 26(2):135-140.
- Willats, J. 1997. Children's Drawing Development. In *Art and Representation: New Principles in the Analysis of Pictures*. Princeton, NJ: Princeton University Press, 287-319.