Human Facial Illustrations: Creation and Psychophysical Evaluation

BRUCE GOOCH
Northwestern University
ERIK REINHARD
University of Central Florida
and
AMY GOOCH
Northwestern University

We present a method for creating black-and-white illustrations from photographs of human faces. In addition an interactive technique is demonstrated for deforming these black-and-white facial illustrations to create caricatures which highlight and exaggerate representative facial features. We evaluate the effectiveness of the resulting images through psychophysical studies to assess accuracy and speed in both recognition and learning tasks. These studies show that the facial illustrations and caricatures generated using our techniques are as effective as photographs in recognition tasks. For the learning task we find that illustrations are learned two times faster than photographs and caricatures are learned one and a half times faster than photographs. Because our techniques produce images that are effective at communicating complex information, they are useful in a number of potential applications, ranging from entertainment and education to low bandwidth telecommunications and psychology research.

Categories and Subject Descriptors: I.3.3 [**Computer Graphics**]: Picture/image Generation—*bitmap and framebuffer operations*; I.3.8 [**Computer Graphics**]: Applications; I.4.3 [**Image Processing and Computer Vision**]: Enhancement—*filtering*

General Terms: Algorithms, Human factors

Additional Key Words and Phrases: Caricatures, Super-portraits, Validation

1. INTRODUCTION

In many applications a non-photorealistic rendering (NPR) has advantages over a photorealistic image. NPR images may convey information better by: omitting extraneous detail; focusing attention on relevant features; and by clarifying, simplifying, and disambiguating shape. The control of detail in an image for purposes of communication is becoming the hallmark of NPR images. This control of image detail is often combined with stylization to evoke the impression of complexity in an image without explicit representation.

This work was carried out under NSF grants NSF/STC for computer graphics EIA 8920219, NSF 99-77218, NSF 99-78099, NSF/ACR, NSF/MRI, and by the DOE AVTC/VIEWS.
Authors' addresses: B. Gooch and A. Gooch, Department of Computer Science, Northwestern University, 1890 Maple, Suite 300, Evanston, IL 60201; email: {bgooch, amygooch}@cs.northwestern.edu; E. Reinhard, School of Computer Science, University of Central Florida, Orlando, FL 32816-2362; email: reinhard@cs.ucf.edu.

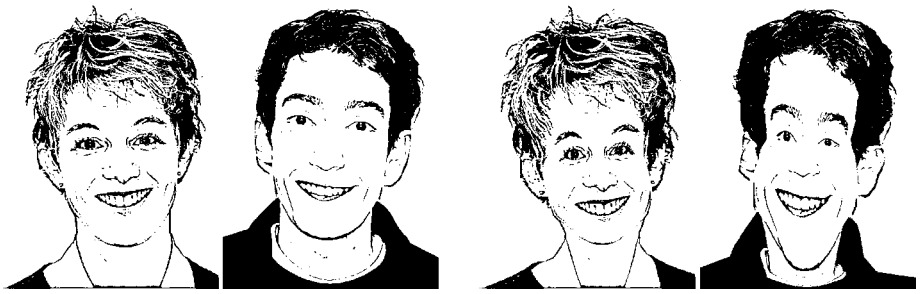

Fig. 1. Left Pair: Examples of illustrations generated using our software. Right Pair: Caricatures created by exaggerating the portraits' representative facial features. This technique is based on the methods of caricature artist Lenn Redman.

Following this trend, we present two techniques that may be applied in sequence to transform a photograph of a human face into a caricature. First, we remove extraneous information from the photograph while leaving intact the lines and shapes that would be drawn by cartoonists. The output of the first step is a two-tone image that we call an *illustration*. Second, this facial illustration may then be warped into a caricature using a practical interactive technique developed from observations of a method used to train caricature artists. Examples of images produced with our algorithms are shown in Figure 1. Compared to previous caricature generation algorithms this method requires only a small amount of untrained user intervention.

As with all non-photorealistic renderings, our images are intended to be used in specific tasks. While we provide ample examples of our results throughout the paper, relying on visual inspection to determine that one method out-performs another, is a doubtful practice. Instead, we show in this paper that our algorithms produce images that allow observers to perform certain tasks better with our illustrations and caricatures than with the photographs they were derived from.

Measuring the ability and effectiveness of an image to communicate the intent of its creator can only be achieved in an indirect way. In general, a user study is conducted whereby participants perform specific tasks on sets of visual stimuli. Relative task performance is then related to the images' effectiveness. If participants are statistically better at performing such tasks for certain types of images, these image types are said to be better at communicating their intent for the given task.

For the algorithms presented in this article, we are interested in the speed and accuracy with which human portraits can be recognized. In addition, we are interested in the speed and accuracy with which human portraits may be learned. If it can be shown that our illustrations afford equal or better performance in these tasks over photographs, then the validity of our algorithms is demonstrated and many interesting applications are opened up. We conducted such a validation experiment, and our results show that recognition speed and accuracy as well as learning accuracy are unaffected by our algorithms, and that learning speed is in fact improved.

This makes our techniques suitable for a wide variety of applications. We provide three examples of potential applications: First, the resulting images are suitable for rapid transmission over low bandwidth networks. Two-tone images may be stored with only one bit per pixel and image compression may further reduce memory requirements. As such, we foresee applications in telecommunications where users may transmit illustrated signature images of themselves in lieu of a caller I.D. number when making phone calls or when using their PDA's. While wireless hand-held digital devices are already quite advanced, bandwidth for rapidly transmitting images is still a bottleneck.

Second, because learning tasks can be sped up by using our images, visual learning applications may benefit from the use of images created using our process. We envision lightweight everyday applications

that could, for example, allow a guest instructor to lean the names of all the students in a class, or allow firemen to learn the names and faces of all the residents while en route.

Third, research in face recognition may benefit from using our techniques. The learning speedup and the recognition invariance demonstrated in our user study suggests that different brain structures or processing strategies may be involved in the perception of artistic images than in the perception of photographs.

In the remainder of this paper we show how illustrations are derived from photographs (Section 2) and how caricatures are created from illustrations (Section 3). These two steps are then subjected to a user study (Section 4), leading to our conclusions (Section 5).

2. ILLUSTRATIONS

Previous research has shown that black-and-white imagery is useful for communicating complex information in a comprehensible and effective manner while consuming less storage [Ostromoukhov 1999; Salisbury et al. 1997, 1996, 1994; Winkenbach and Salesin 1994; Tanaka and Ohnishi 1997]. With this idea in mind, we would like to produce easily recognizable black-and-white illustrations of faces. Some parts of the image may be filled in if this increases recognizability. However, shading effects that occur as a result of the lighting conditions under which the photograph was taken should be removed because they are not an intrinsic aspect of the face. In addition, we would like to be able to derive such illustrations from photographs without skilled user input.

Creating a black-and-white illustration from a photograph can be done in many ways. A number of proposed methods are stroke-based and rely heavily on user input [Durand et al. 2001; Ostromoukhov 1999; Sousa and Buchanan 1999; Wong 1999]. In addition, stroke-based methods are mainly concerned with determining stroke placement in order to maintain tonal values across an object's surface. For our application, we prefer a method that could largely be automated and does away with tonal information.

One second possible method for creating facial illustrations is to only draw pixels in the image with a high intensity gradient [Pearson and Robinson 1985; Pearson et al. 1990]. These pixels can be identified using a valley filter and then thresholded by computing the average brightness and setting all pixels that are above average brightness to white and all other pixels to black. However, this approach fails to preserve important high luminance details, and thus only captures some facial features. While the resulting image can be interpreted as a black-and-white drawing, we find that it leaves "holes" in areas that should be all dark and that filling in the dark parts produces images that are more suitable. This filling in can be accomplished by thresholding the input luminances separately and multiplying the result of this operation with the thresholded brightness image [Pearson and Robinson 1985].

A third approach could be the use of edge detection algorithms to remove redundant data. These algorithms are often used in machine vision applications. Most of these edge detection algorithms produce thin lines that are connected. While connectedness is a basic requirement in machine vision, it is specifically not needed for portraits of faces and may reduce recognizability [Davies et al. 1978].

To comply with our requirements that the method needs minimal trained user input and produces easily recognizable images, we choose to base our algorithm on a model of human brightness perception. Such models are good candidates for further exploration because they flag areas of the image where interesting transitions occur, while removing regions of constant gradient. How this is achieved is explained next. In addition, we preserve a sense of absolute luminance levels by thresholding the input image and adding this to the result. The general approach of our method is outlined in Figure 2.

2.1 Contrast and Brightness Perception

Before introducing our algorithm for computing illustrations, in this subsection, we briefly summarize models of brightness perception. It should be noted that there are many reasonable brightness

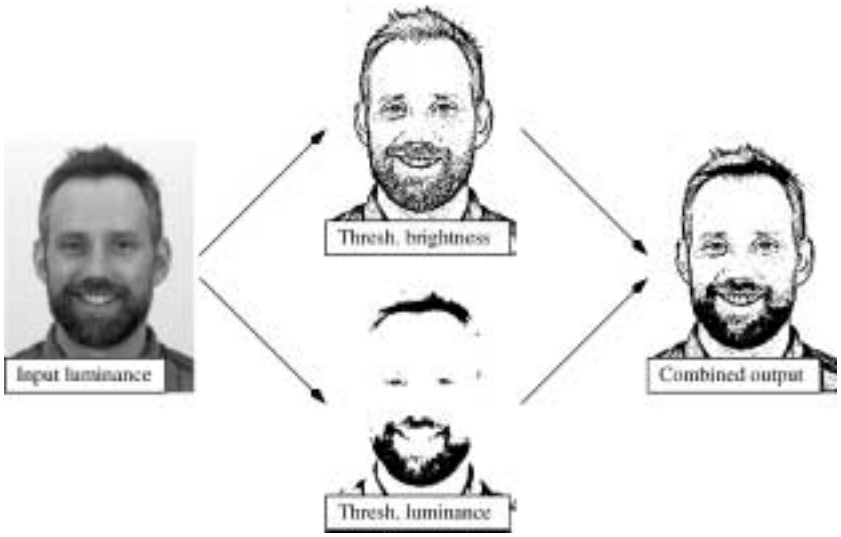

Fig. 2. Brightness is computed from a photograph, then thresholded and multiplied with thresholded luminance to create a line art portrait.

models available and that this is still an active area of research, which is much more fully described in excellent authoritative texts such as Boff et al. [1986], Gilchrist [1994], Wandell [1995], and Palmer [1999].

While light is necessary to convey information from objects to the retina, the human visual system attempts to discard certain properties of light [Atick and Redlich 1992; Blommaert and Martens 1990]. An example is brightness constancy, where brightness is defined as a measure of how humans perceive luminance [Palmer 1999].

Brightness perception can be modeled using operators such as differentiation, integration and thresholding [Arend and Goldstein 1987; Land and McCann 1971]. These methods model lateral inhibition which is one of the most pervasive structures in the visual nervous system [Palmer 1999]. Lateral inhibition is implemented by a cell's receptive field having a center-surround organization. Thus cells in the earliest stages of human vision respond most vigorously to a pattern of light which is bright in the center of the cell's receptive field and dark in the surround, or vice-versa. Such antagonistic center-surround behavior can be modeled using neural networks, or by computational models such as Difference of Gaussians [Blommaert and Martens 1990; Cohen and Grossberg 1984; Gove et al. 1995; Hansen et al. 2000; Pessoa et al. 1995], Gaussian smoothed Laplacians [Marr and Hildreth 1980; Marr 1982] and Gabor filters [Jernigan and McLean 1992].

Closely related to brightness models are edge detection algorithms that are based on the physiology of the mammalian visual system. An example is Marr and Hildreth's zero-crossing algorithm [Marr and Hildreth 1980]. This algorithm computes the Laplacian of a Gaussian blurred image (LoG) and detects zero crossings in the result. The LoG is a two dimensional isotropic measure of the second spatial derivative of an image. It highlights regions of rapid intensity change and can therefore be used for edge detection. Note that the Laplacian of Gaussian can be closely approximated by computing the difference of two Gaussian blurred images, provided the Gaussians are scaled by a factor of 1.6 with respect to each other [Marr 1982]; a feature also employed in our computational model.

2.2 Computing Illustrations

To create illustrations from photographs we adapt Blommaert and Martens' [1990] model of human brightness perception, which was recently shown to be effective in a different type of application (tone reproduction, see Reinhard et al. [2002]). The aim of the Blommaert model is to understand brightness perception in terms of cell properties and neural structures. For example, the scale invariance property of the human visual system can be modeled by assuming that the outside world is interpreted at different levels of resolution, controlled by varying receptive field sizes. Blommaert and Martens [1990] demonstrate that, to a first approximation, the receptive fields of the human visual system are isotropic with respect to brightness perception, and can be modeled by circularly symmetric Gaussian profiles R_i:

$$R_i(x, y, s) = \frac{1}{\pi(\alpha_i s)^2} \exp\left(-\frac{x^2 + y^2}{(\alpha_i s)^2}\right). \tag{1}$$

These Gaussian profiles operate at different scales s and at different image positions (x, y). We use R_1 for the center and R_2 to model the surround and let $\alpha_1 = 1/(2\sqrt{2})$. The latter ensures that two standard deviations of the Gaussian overlap with the number of pixels specified by s. For the surround we specify $\alpha_2 = 1.6\alpha_1$. A neural response V_i as function of image location, scale and luminance distribution L can be computed by convolution:

$$V_i(x, y, s) = L(x, y) \otimes R_i(x, y, s). \tag{2}$$

The firing frequency evoked across scales by a luminance distribution L is modeled by a center-surround mechanism:

$$V(x, y, s) = \frac{V_1(x, y, s) - V_2(x, y, s)}{2^\phi/s^2 + V_1(x, y, s)}, \tag{3}$$

where center V_1 and surround V_2 responses are derived from Eqs. (1) and (2). Subtracting V_1 and V_2 leads to a Mexican hat shape, which is normalized by V_1. The term $2^\phi/s^2$ is introduced to avoid the singularity that occurs when V_1 approaches zero and models the (scale-dependent) rest activity associated with the center of the receptive field [Blommaert and Martens 1990]. The value 2^ϕ is the transition flux at which a cell starts to be photopically adapted.

In the Blommaert model, the parameter 2^ϕ is set to 100 cd arcmin^2 m^{-2}. Because in our application we deal with low dynamic range images, as well as an uncontrolled display environment (see below), we adapt the model heuristically by setting $\phi = 1$. We have found that this parameter can be varied to manipulate the amount of fine detail present in the illustration. An expression for brightness B is now derived by summing V over all scales:

$$B(x, y) = \sum_{s=s_0}^{s_{max}} V(x, y, s). \tag{4}$$

The Blommaert model, in line with other models of brightness perception, specifies these boundaries in visual angles, ranging from 2 arcmin to 50 degrees. In a practical application, the size of each pixel as it is displayed on a monitor, is generally unknown. In addition, the distance of the viewer can not be accurately controlled. For these reasons, we translate these angles into image sizes under the reasonable assumption that the smallest size is 1 pixel ($s_0 = 1$). The number of discrete scales is chosen to be 8, which provides a good trade-off between speed of computation and accuracy of the result. These two parameters fix the upper boundary s_{max} to be $1.6^8 \approx 43$ pixels. For computational convenience, the scales s are spaced by a factor of 1.6 with respect to each other. This allows us to reuse the surround computation at scale s_i for the center at scale s_{i+1}.

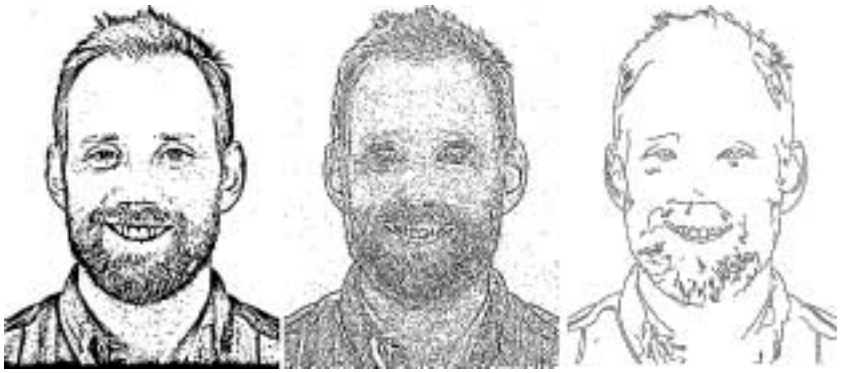

Fig. 3. Thresholded brightness (left), high-pass filtered image followed by thresholding (middle) and Canny's edge detector (right).

The result of these computations is an image which could be seen as an interpretation of human brightness perception. One of the effects of this computation is that constant regions in the input image remain constant in the brightness image. Also, areas with a constant gradient are removed and become areas of constant intensity. In practice, this has the effect of removing shading from an image. Removing shading is an advantage for computing illustrations because shading information that is typically not shown in illustrations. Brightness images are typically grey with lighter and darker areas near regions with nonzero second derivatives. In illustrations, these regions are usually indicated with lines. There is therefore a direct relation between the information present in a brightness image and the lines drawn by illustrators.

As such, the final step is converting our brightness representation into a two-tone image that resembles an illustration. We do this by computing the average intensity of the brightness image and setting all pixels that are above this threshold to white and all other pixels to black. Dependent on the composition of the photograph, this threshold may be manually increased or decreased. However, it is our experience that the average brightness is always a good initial estimate and that deviations in choice of threshold tend to be small. All other constants in the brightness model are fixed as indicated above and therefore do not require further human intervention.

Figure 3 shows the result of thresholding our brightness representation and compares it to thresholding a high pass filtered image (middle) and an edge detected image (right). Thresholding a high-pass filtered image and edge detecting are perhaps more obvious strategies that could potentially lead to similar illustrations. Figure 3 shows that this is not necessarily the case. The high-pass filtered image was obtained by applying a 5×5 Gaussian kernel to the input image and subtracting the result from the input image. The threshold-level was chosen to maximize detail while at the same time minimizing salt-and-pepper noise. Note that it is difficult to simultaneously achieve both goals within this scheme. The comparison with Canny's edge detector [Canny 1986] is provided to illustrate the fact that connected thin lines are less useful for this particular application.

The thresholded brightness image can be interpreted as a black-and-white illustration, although we find that filling in the dark parts produces images that are easier to recognize. Filling in is accomplished by thresholding the luminance of the input image separately and multiplying the result of this operation with the thresholded brightness image [Pearson and Robinson 1985]. The threshold value is chosen manually according to taste, but often falls in the range from about 3 to 5% grey. The process of computing a portrait is illustrated in Figure 2.

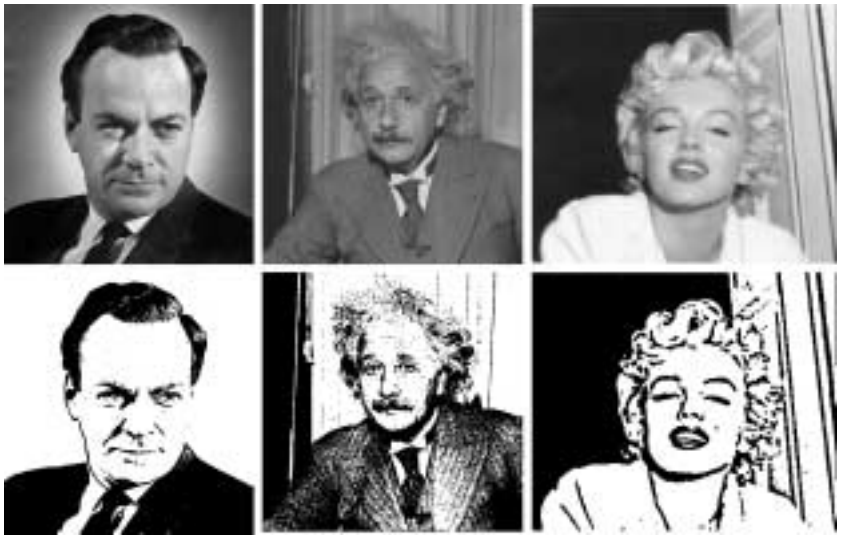

Fig. 4. Source images (top) and results of our perception based portrait algorithm (bottom).

Table I. Storage Space (in Bits Per Pixel) for Photographs and Facial Illustrations Computed Using Our Method

Size (pixels)	429×619	320×240	160×160
Photograph	1.21	0.96	1.20
Illustration	0.10	0.11	0.19

Our facial illustrations are based on photographs but contain much less information, as shown in Figure 4. For example, shading is removed from the image, which is a property of Difference of Gaussians approaches. As such, the storage space required for these illustrations is decreased from the space needed to store photographs. See Table I.

On a 400-MHz R12k processor, the computation time for a 1024^2 image is 28.5 s, while a 512^2 can be computed in 6.0 s. These timings are largely due to the FFT computation used to compute the convolution of Eq. (2). We anticipate that these images could be computed faster with approximate methods, although this could be at the cost of some quality. In particular, we believe that a multiresolution spline method may yield satisfactory results [Burt and Adelson 1983].

On the other hand, it should be pointed out that the brightness computation involves a number of FFT computations to facilitate the convolution operations. This makes the algorithm relatively expensive. Also, the brightness threshold and the threshold on luminance of the source image are currently specified by hand. While a reasonable initial value may be specified based on the average intensity found in the brightness images, a small deviation from this initial threshold almost always improves the result. Further research may automate this process and so make this method better suitable for producing animated illustrations from video sequences. Should better models of brightness perception become available, then these may improve our results. In particular, because human visual perception is to a lesser extent sensitive to absolute intensity levels, a brightness model that incorporates both relative as well as absolute light levels may further improve our results. Although Blommaert and Martens discuss a notion of absolute light levels [Blommaert and Martens 1990], for our application their model requires the additional application of thresholded absolute luminance levels. It would be better if this could be incorporated directly into the brightness model.

Fig. 5. Left Trio: Photographic example. Right Trio: Facial illustration example. In both examples the first images are of 50% anti-caricatures, the second images are the source images, and the third images are 50% super portraits.

While Figure 4 allows the reader to subjectively assess the performance of our algorithm, its real merit lies in the fact that specific tasks can be performed quicker using facial illustration images than when using photographs. This remarkable finding is presented in Section 4.

Finally, some of the facial features that a cartoonist would draw, are absent from our illustrations while some noise is present. Parameters in our model may be adjusted to add more lines, but this also increases the amount of noise in the illustrations. While our algorithm produces plausible results with the parameters outlined above, future research into alternative algorithms may well lead to improved quality.

2.3 Creating a Super-Portrait

Super-portraits are closely related to the *peak shift* effect, which is a well-known principle in animal learning [Hansen 1959; Ramachandran and Hirstein 1999]. It is best explained by an example. Suppose a laboratory rat is taught to discriminate a square from a rectangle by being rewarded for choosing the rectangle, it will soon learn to respond more frequently to the rectangle. Moreover, if a prototype rectangle of aspect ratio 3:2 is used to train the rat, it will respond even *more* positively to a longer and thinner rectangle with an aspect ratio of 4:1. This result implies the rat is not learning to value a particular rectangle but a *rule*, in this case that rectangles are better than squares. So the more oblong the rectangle, the better the rectangle appears to the rat.

Super-portraits of human faces are a well studied example of the peak shift effect in human visual perception [Benson and Perret 1994; Rhodes et al. 1987; Rhodes and Tremewan 1994, 1996; Stevenage 1995]. It has been postulated that humans recognize faces based on the amount that facial features deviate from an average face [Tversky and Baratz 1985; Valentine 1991]. Thus, to produce a super portrait, features are exaggerated based on how far the face's features deviate from an average or norm face [Brennan 1985]. Figure 5 shows examples of super-portraits.

Two paradigms exist that explain how humans perform face recognition. In the average-based coding theory, a "feature space distance" from an average face to a given face is encoded by the brain [Valentine 1991]. An alternative model of face recognition is based on exemplars [Lewis and Johnston 1998], where face representations are stored in memory as absolutes. Both models equally account for the effects observed in face recognition tasks, but the average-based coding paradigm lets itself be cast more easily into a computational model.

In our approach, super-portraits may be created in a semiautomatic way. A face is first framed with four lines that comprise a rectangle, as shown in Figure 7. Four vertical lines are then introduced marking the inner and outer corners of the eyes, respectively. Next, three additional horizontal lines mark the position of the eyes, the tip of the nose, and the mouth. We call this set of horizontal and vertical lines a facial feature grid (FFG).

To generate a FFG for a norm face, we apply a previously defined metric [Redman 1984]. The vertical lines are set to be equidistant, while the horizontal eye, nose and mouth lines are assigned distance values 4/9, 6/9 and 7/9 respectively, from the top of the frame. The norm FFG is automatically computed when the face is framed by the user. Gridding rules can also be specified for profile views [Redman 1984], but for the purpose of this work, we have constrained our input to frontal views.

When a feature grid is specified for a given photograph or portrait, it is unlikely to coincide with the norm FFG. The difference between the norm face grid and the user-set grid can be exaggerated by computing the vectors between corresponding vertices in both grids. Then, these vectors are scaled by a given percentage and the source image is warped correspondingly. When this percentage is positive, the result is called a super portrait, whereas negative percentages give rise to anticaricatures, images that are closer to the norm face than the input image (Figure 5).

3. CARICATURES

The documented ability of caricatures to augment the communication content of images of human faces motivated the investigation of computer generated caricatures [Brennan 1985; Benson and Perret 1991; Rhodes et al. 1987; Stevenage 1995]. To create a caricature, the difference between an average face and a particular face can be computed for various facial features, which is then exaggerated by a specified amount. We describe methods for automatically deviating from an average face, as well as techniques that allow meaningful warping to perform more extreme caricaturing.

Traditionally caricatures have been created by skilled artists who use lines to represent facial features. The skill of the artist lies in knowing which particular facial features are essential and which are incidental. For facial caricatures both artists and psychologists agree that the feature shift for a particular face should exaggerate the differences from an average face [Brennan 1985; Redman 1984; Rhodes et al. 1987]. Automatically creating such drawings has been an elusive goal, and attempts to automate this process are sparse.

The most well-known attempt is the “Caricature Generator” [Brennan 1985], which is based on the notion of an average face. The positions of 165 feature points are indicated by a knowledgeable user marking points on a scanned photograph with mouse clicks. The points for the given face are then compared with the positions of similar points on an average face. By moving the user defined points away from the average, an exaggerated effect can be created. A line drawing is created by connecting the feature points with lines. A caricature is created by translating the feature points over some distance and then connecting them with lines. This method was later extended to allow the feature translation to be applied to the input image in order to produce a photographic caricature [Benson and Perret 1991]. The Caricature Generator is used in many psychophysical experiments and has become a *de facto* standard for conducting research in face recognition [Benson and Perret 1994; Rhodes et al. 1987; Rhodes and Tremewan 1994, 1996; Stevenage 1995] (see also Section 4). Examples of facial illustrations and caricatures created using an implementation of the “Caricature Generator” software are shown in Figure 6.

A second semi-automated caricature generator is based on simplicial complexes [Akleman et al. 2000]. The deformations applied to a photograph of a face are defined by pairs of simplices (triangles in this case). Each pair of triangles specifies a deformation, and deformations can be blended for more general warps. This system is capable of interactively producing extreme exaggerations of facial features, but requires experience to meaningfully specify source and target simplices.

Both previous methods require expert knowledge or skilled user input, which limits their applicability for every-day use. We propose semi-automatic methods to produce super-portraits and caricatures which rely much less on the presence of well-trained users.

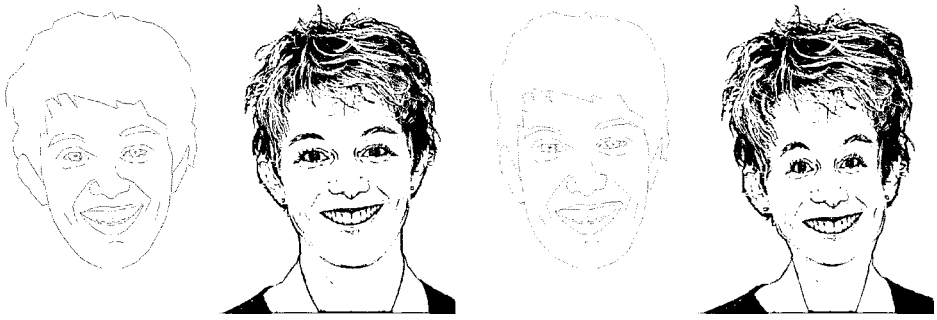

Fig. 6. Left Pair: An example of a line art image created using an implementation of the "Caricature Generator" software compared to a facial illustration created using our technique. Right Pair: An example of a caricature created using an implementation of the "Caricature Generator" software compared to a caricature created using our technique.

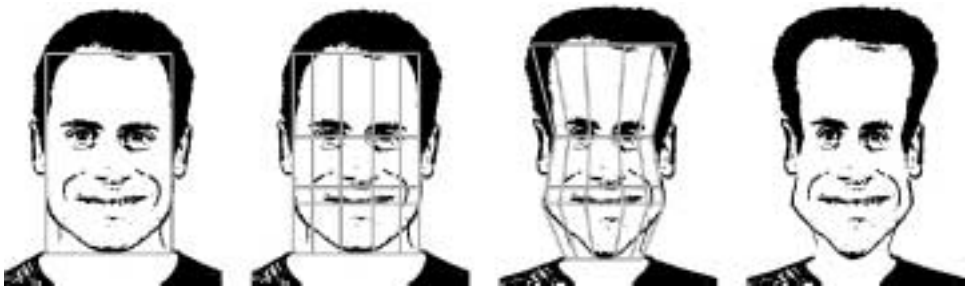

Fig. 7. First: The face is framed by four border lines. Second: Facial features and interior lines are matched. Third: Both grid and underlying image are warped interactively. Fourth: The resulting caricature.

3.1 Creating a Caricature

As a manifestation of the peak shift effect, super-portraits are useful in the study of human perception. However, more extreme distortions may be required in other applications. We therefore extend our algorithm to allow more expressive freedom. Based on the feature grid as described above, we allow vertices on the left and right of the grid to be manipulated individually. In addition, grid lines may be moved. This user interaction is interactively visualized by warping the image according to the position of the vertices (Figure 7). This process is constrained by disallowing the selection and manipulation of internal vertices. We believe that the resulting system is flexible enough to create amusing caricatures, while at the same time protecting untrained users from producing unrecognizable faces. Figure 8 compares examples of caricatures drawn by noted caricature artist Ric Machin to images created using our system. These examples demonstrate that by warping faces along a feature grid exaggerations similar to those used by professional artists can be achieved. The implementation is straightforward, both in OpenGL and Java, and interactive manipulation was achieved on current workstations. Caricatures created by users who were given minimal verbal training are presented in Figure 9. Within two minutes, users were able to create a caricature. Often they remarked that they were able to create several evocative caricatures from just a single portrait.

4. VALIDATION

The effectiveness of the portraits and caricatures is evaluated for specific tasks by means of a psychophysical study. We are interested in the influence of our facial illustration and caricature algorithms

Fig. 8. The exaggeration of facial features used by a professional caricature artist can easily be obtained by a novice user with our algorithm. First and Third: Examples of caricatures by noted caricature artist Ric Machin. Second and Fourth: Examples of caricatures created using our interactive software.

Fig. 9. Examples of caricatures created by novice users of the software, who were given minimal, one to two minutes, verbal instruction in the use of the software. They were able to produce the resulting caricature images in periods ranging from one to four minutes. The top row shows photographs used to inspire the caricatures in the columns below each of them.

on recognition and learning tasks. In particular, we assess the speed and accuracy of these two tasks. The first hypothesis is that if either algorithm does not affect the recognition of familiar faces with respect to photographs, then the information reduction afforded by these algorithms is relatively benign and the resulting images can be substituted in tasks where recognition speed is paramount.

Second, we ask the question if our portrait and caricature algorithms affect the speed and accuracy of learning. Past experiments have shown that learning in simplified environments may proceed faster

than for similar learning tasks executed in the full environment [Brennan 1985; Rhodes et al. 1987]. As our illustrations and caricatures may be regarded as simplified with respect to the input photographs, learning them may be easier than learning the associated photographs.

To test these hypotheses, two experiments were performed that are replications of earlier distinctiveness experiments [Stevenage 1995]. While these previous experiments assessed the effect of human drawn portraits and caricatures on recognition and learning speed, these same experiments are used here to validate the computer-generated illustrations and caricaturing techniques. In addition, the computer-generated illustrations and caricatures are compared with the base photographs in terms of recognition and learning speeds.

4.1 Recognition Time

This experiment assesses the recognition time of familiar faces presented as illustrations, caricatures, and photographs. Based on the results obtained with Caricature Generator images [Brennan 1985; Benson and Perret 1991; Rhodes et al. 1987], we expect that photographs may be recognized faster than both facial illustrations and caricatures, while caricatures would elicit a faster response time than the facial illustrations [Benson and Perret 1991].

Subjects were presented with sequences of images of familiar faces. Each participant was asked to say the name of the person pictured as soon as that person's face was recognized. Reaction times as well as accuracy of the answers were recorded. Images were presented to the participants in three separate conditions, using either photographs and facial illustrations, photographs and caricatures, or facial illustrations and caricatures. The details of this experiment are presented in Appendix A.

The difference between the mean recognition time for photographs and illustrations of familiar faces was not statistically different, hovering around 1.89 seconds. The caricatures were recognized on average 0.09 seconds slower than photographs. There was no statistical difference in the reaction time between caricatures and facial illustrations. In each condition, the accuracy of recognition was higher than 98%, indicating that there is no speed for accuracy trade-off in this experiment.

The recognition accuracy measured in our experiment is compared with studies using hand-drawn images and images created using the Caricature Generator [Brennan 1985] in Table IV. The accuracy witnessed with our algorithms is markedly better than the results obtained with the Caricature Generator, while at the same time leading to shorter response times (compare with Rhodes et al. [1987]). The latter produces images that are made up of a sparse set of thin lines (Figure 6), which may help explain the differences [Davies et al. 1978].

We conclude that substituting illustrations for fully detailed photographs has no significant recognition speed cost and can therefore be used for tasks in which speed of recognition is central. Caricatures cause a slight degradation in reaction time. However, half of the participants laughed out loud during this experiment when shown caricatures of familiar faces. We take this to mean that caricatures could be used as intended for entertainment value in situations where recognition speed is not of the utmost importance.

4.2 Learning Speed and Accuracy

At least one study indicates that caricatures do not improve the ability to learn [Hagen and Perkins 1983], while others have shown that caricatures of unfamiliar faces can actually be learned quicker than the same faces shown as fully detailed drawings or as photographs [Mauro and Kubovy 1992; Stevenage 1995]. We hypothesize that the outcome of these studies is strongly dependent on the particular techniques used to create the stimuli. Because our approach to creating illustrations and caricatures is very different from those used in previous user studies, we have subjected participants to a learning task to assess how our illustrations and caricatures influence the ability to learn human faces.

In this experiment, each participant was presented with images of twelve unfamiliar faces in sequence. Each face was verbally assigned a name. Each subject was shown exclusively either photographs, facial illustrations, or caricatures, but never a combination of modes. Next, the images were reordered and presented again. The participant was asked to recall the name corresponding to each image. During each iteration of this process, the experimenter corrected mistakes and repeated the names that the participant could not remember. The images were then shuffled, and the process was repeated until the subject could correctly name all twelve faces once without error. The details of this experiment are presented in Appendix B.

Learning to recognize twelve unfamiliar faces was accomplished more than twice as fast in trials with illustrations compared to trials with photographs. Caricatures were learned over 1.5 times faster than photographs.

This experiment was followed by a test for learning accuracy. Participants who were trained using the portrait or caricature images participated in a follow-up experiment using the original photographs in an otherwise identical set-up. In this experiment, the number of incorrectly named faces was recorded. The training using either caricatures or facial illustrations both resulted in a 98% naming accuracy for photographs. Details are given in Appendix C.

Thus, it appears that the rate of learning can be increased using images generated with the line art algorithms without suffering a decrease in accuracy. For this effect to occur, the training data has to be simpler to learn than the test data, while still retaining all relevant cues [Sweller 1972].

5. CONCLUSIONS AND FUTURE WORK

Because humans appear to employ dedicated processing for face recognition [Biederman and Kalocsai 1998; Ellis 1986; Gauthier et al. 1999], NPR algorithms for image processing on portraits need to be designed with care. Certain features need to remain present in the image to preserve recognizability. Our method for producing illustrations from photographs achieves this by applying a modified model of human brightness perception to photographs. The resulting illustrations are as easy to recognize as photographs, and are faster to learn.

The second level of processing that we presented, is an image warping technique that produces super-portraits as well as caricatures. Super-portraits are essentially the application of the peak shift effect to human face perception. Our method of producing these constitutes a viable alternative for further research into this phenomenon, and may prove to be a good replacement for the Caricature Generator that is regarded as the de facto standard.

The caricatures that we produce have largely similar characteristics in terms of recognition and learning tasks as the illustrations we derive from photographs, and can therefore be applied in entertainment oriented applications without significantly impeding task performance.

The learning speedup and the recognition invariance demonstrated in the psychophysical experiments suggests that different brain structures may be involved in the perception of artistic images than are involved in the perception of photographs. Therefore, it may be interesting to conduct these same psychophysical experiments again as experimental probes in functional magnetic resonance imaging (fMRI) experiments. The fMRI experiments could be used in an attempt to decipher the differences between human perception of art and photographs.

Transfer of information in learning tasks is another area for future investigation. For example, once the front view of a face is learned, how well can a three quarter view or profile be recognized? Could a facial representation learned in a line or caricature format be recognized in a photo of a crowded room or in person? The answers to these questions may be significant, depending on the application.

There are many more possible future directions for this work. Animating human facial illustrations is a natural extension of this work. Processing a series of images using our techniques would lead to an

animated portrait. Currently we use an image filtering technique to produce facial illustrations. Another possibility is to incorporate some of the the stroke based techniques in order to produce images more reminiscent of drawings. Also, the conditions under which photographs are taken, may influence the output of our illustrations algorithm. For example skin blemishes, freckles, or even a slight beard may sometimes lead to artifacts in the illustrations. Although this issue can be circumvented by making the subjects wear makeup, or even by touching up the photographs in a drawing program, a more principled approach would be to further increase the robustness of our algorithm against these factors.

APPENDIXES

A. RECOGNITION SPEED

Subjects. 42 graduate students, postgraduates and research staff acted as volunteers.

Materials. We used 60 images depicting the faces of 20 colleagues of the volunteers. Each face was depicted as a grey-scale photograph, an illustration, and a caricature. The photographs were taken indoors using a Kodak 330 digital camera with the flash enabled. In a pilot study five independent judges rated each illustration and caricature as a good likeness of the face it portrayed. The images were displayed on a Dell Trinitron monitor at a distance of 24 inches. The monitor's background was set to black and displayed images subtending a visual angle of 12.9 degrees. Images were shown for 5 seconds at 5-second intervals.

Procedure. We conducted 3 two-part experiments, each with 14 participants. The first part allowed participants to rate their familiarity with a list of 20 names on a 7-point scale with a purpose designed user interface. Participants were given the following written instructions: "Please read each name and form a mental image of that person's face. Then say the name aloud. Finally, rate the accuracy of your mental image of that person and position the slider accordingly. Please repeat this for each person on the list." By pronouncing the names of the people that were rated, participants tend to reduce the "tip-of-the-tongue" effect where a face is recognized without being able to quickly recall the associated name [Stevenage 1995; Yarmey 1973; Young et al. 1985].

In the second part of this experiment, the 12 highest rated faces are selected for each participant and were shown in 2 of 3 possible conditions. Participants in Experiment A.1 saw photographs and facial illustrations. Experiment A.2 consisted of photographs and caricatures, and Experiment A.3 consisted of facial illustrations and caricatures. The written instructions for this part were: "In this experiment you will be shown pictures of people's faces you may know. Each picture will be shown for five seconds followed by a 5-second interval. Please say the name of each person as soon as you recognize this person." The experimenter provided additional verbal instructions to reduce the surprise associated with showing the first image (a practice trial), and to further reduce the tip-of-the-tongue effect, participants were told that first, last or both names could be given, whichever was easiest. One experimenter recorded the accuracy of the answers and the response time for each image was recorded by a second experimenter who pressed a key at the end of the response. This stopped the timer that was started automatically upon display of the image.

Results. Subjects were faster at naming photographs ($M = 1.89$s) compared to caricatures ($M = 2.01$s, $p < 0.01$). There was no difference between the time to name photos compared with illustrations ($p = 0.55$) and a marginal advantage for naming facial illustrations compared to caricatures ($p = 0.07$). The accuracy for recognizing photos, facial illustrations and caricatures are 98%, 99% and 98% respectively. Table II provides minimum, maximum, and mean times recorded for each condition on each experiment.

Table II. Recognition Speed Results, Showing the Minimum, Maximum, and Mean Time Over Average Subject Data for Each Condition in Each Experiment

Condition	Min	Max	Mean	Std. Error
	Experiment A.1 ($p = 0.011$)			
Photograph	1.53 s	2.34 s	1.89 s	0.080
Caricature	1.57 s	2.57 s	2.01 s	0.094
	Experiment A.2 ($p = 0.072$)			
Portrait	1.47 s	2.62 s	1.20 s	0.089
Caricature	1.47 s	2.83 s	2.11 s	0.120
	Experiment A.3 ($p = 0.555$)			
Photograph	1.38 s	2.30 s	1.85 s	0.069
Portrait	1.51 s	2.32 s	1.85 s	0.096

Table III. Learning Speed Experiments, Showing the Minimum, Maximum, and Mean Number of Trial Iterations for the Experiment Presented in Appendix B

	Trials			
Condition	Min	Max	Mean	Std. Error
Photographs	1	8	5.4	0.79
Facial Illustrations	1	4	2.3	0.26
Caricatures	1	7	3.5	0.58

B. LEARNING SPEED

Subjects. 30 graduate students, postgraduates and research staff acted as volunteers. They were selected for unfamiliarity with the faces presented in this experiment.

Materials. We used grey-scale photographs of the faces of 6 males and 6 females. An identical pilot study as in Experiment A was carried out and the 12 facial illustrations and 12 caricatures derived from these photos were all rated as good likenesses. All photos, facial illustrations and caricatures were printed on a laser printer at a size of $6'' \times 8''$ at 80 dpi and mounted on matting board. Each face was randomly assigned a two-syllable first name from a list of the most popular names of the 1970's (taken from www.cherishedmoments.com/most-popular-baby-names.htm). In a separate pilot study, this set of names was rated for distinctiveness and names causing confusion were replaced.

Procedure. Each participant was given a list with 12 names and then asked to learn to match these names with the 12 faces. The participants were divided into 3 groups of 10 and each participant was individually presented exclusively with either photographs, illustrations or caricatures. Each participant was first shown all 12 faces, one image at a time, for about 3 seconds and was told the name assigned to that face. The faces were then shuffled and individually presented to the subject who was now asked to recall each name. The number of incorrect responses was recorded and the participant was corrected if mistakes were made. This procedure was repeated, shuffling the faces between each trial, until all twelve faces were correctly named in two successive sequences. The number of trials taken to reach this criterion represents the dependent variable in this learning experiment.

Results. The statistics for the rate of learning (number of trials) for each representation of the faces is shown in Table III. Illustrations were learned substantially faster than photos ($p < 0.001$). In this trial caricatures versus photos, and facial illustrations versus caricatures could not be distinguished

Table IV. Comparison of the Recognition Accuracy of Facial Illustrations, Caricatures, and Photographs Across Studies

	Accuracy			Method Details
User study	Portraits	Caricatures	Photographs	Portraits/Caricatures
Stevenage [1995]	96%	100%	–	Traced/Pro. Artist
Rhodes et al. [1987]	38%	33%	–	Skilled user/Autom.
Current study	99%	98%	98%	Autom./Unskilled user

statistically ($p = 0.093$, $p = 0.081$). A preliminary conclusion is that learning appears to be quickest when the faces were presented as illustrations, followed by caricatures and then photographs.

C. LEARNING ACCURACY

Subjects. 20 subjects participating in Experiment B who were presented with either the illustrations or the caricatures.

Materials. Same as in Experiment B.

Procedure. In this experiment we explored whether caricatures and facial illustrations result in a difference in learning accuracy. After participating in Experiment B, subjects were shown 12 photographs in random order and were asked to recall the associated names.

Results. See Table IV. The number of incorrectly identified faces was recored for each participant. The accuracy was 98% for training on illustrations as well as for training on caricatures. Hence, there was no measurable difference in accuracy between participants trained with facial illustrations or caricatures.

ACKNOWLEDGMENTS

We would like to thank Mike Stark, Bill Martin, Peter Shirley, Richard F. Riesenfeld, Bill Thompson, Charles Hansen and the University of Utah Graphics Group for their help in this work.

REFERENCES

AKLEMAN, E., PALMER, J., AND LOGAN, R. 2000. Making extreme caricatures with a new interactive 2D deformation technique with simplicial complexes. In *Proceedings of Visual'2000*.

AREND, L. AND GOLDSTEIN, R. 1987. Lightness models, gradient illusions, and curl. *Percept. Psychophys. 43*, 65–80.

ATICK, J. J. AND REDLICH, N. A. 1992. What does the retina know about natural scenes? *Neur. Comput. 4*, 196–210.

BENSON, P. J. AND PERRET, D. I. 1991. Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology 3*, 1, 105–135.

BENSON, P. J. AND PERRET, D. I. 1994. Visual processing of facial distinctiveness. *Perception 23*, 75–93.

BIEDERMAN, I. AND KALOCSAI, P. 1998. Neural and psychological analysis of object and face recognition. In *Face Recognition: From Theory to Applications*. Springer-Verlag, New York, 3–25.

BLOMMAERT, F. J. J. AND MARTENS, J.-B. 1990. An object-oriented model for brightness perception. *Spatial Vis. 5*, 1, 15–41.

BOFF, K., KAUFMAN, L., AND THOMAS, J. P. 1986. *Handbook of Perception and Human Performance*. Wiley, New York.

BRENNAN, S. E. 1985. Caricature generator: The dynamic exaggeration of faces by computer. *Leonardo 18*, 3, 170–178.

BURT, P. J. AND ADELSON, E. H. 1983. A multiresolution spline with application to image mosaics. *ACM Trans. Graph. 2*, 4, 217–236.

CANNY, J. F. 1986. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intel. 8*, 769–798.

COHEN, M. A. AND GROSSBERG, S. 1984. Neural dynamics of brightness perception: Features, boundaries, diffusion and resonance. *Percept. Psychophy. 36*, 5, 428–456.

DAVIES, G., ELLIS, H., AND SHEPPERD, J. 1978. Face recognition accuracy as a function of mode of representation. *J. Appl. Psych. 63*, 2, 180–187.

DURAND, F., OSTROMOUKHOV, V., MILLER, M., DURANLEAU, F., AND DORSEY, J. 2001. Decoupling strokes and high-level attributes for interactive traditional drawing. In *Rendering Techniques 2001: 12th Eurographics Workshop on Rendering*. Eurographics, 71–82. ISBN 3-211-83709-4.

ELLIS, H. D. 1986. Introduction to aspects of face processing: ten questions in need of answers. In *Aspects of Face Processing*, H. Ellis, M. Jeeves, F. Newcombe, and A. Young, Eds. Nijhoff, Dordrecht, 3–13.

GAUTHIER, I., TARR, M. J., ANDERSON, A. W., SKUDLARSKI, P., AND GORE, J. C. 1999. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat. Neurosci. 2*, 6 (June), 568–573.

GILCHRIST, A. L. ED. 1994. *Lightness, Brightness, and Transparency*. Erlbaum, Mahwah, N.J.

GOVE, A., GROSSBERG, S., AND MINGOLLA, E. 1995. Brightness perception, illusory contours, and corticogeniculate feedback. *Vis. Neurosci. 12*, 1027–1052.

HAGEN, M. A. AND PERKINS, D. 1983. A refutation of the hypothesis of the superfidelity of caricatures relative to photographs. *Perception 12*, 55–61.

HANSEN, M. H. 1959. Effects of discrimination training on stimulus generalization. *J. Exper. Psych. 58*, 3221–3334.

HANSEN, T., BARATOFF, G., AND NEUMANN, H. 2000. A simple cell model with dominating opponent inhibition for robust contrast detection. *Kognitionswissenschaft 9*, 93–100.

JERNIGAN, M. E. AND MCLEAN, G. F. 1992. Lateral inhibition and image processing. In *Non-linear vision: Determination of neural receptive fields, function, and networks*, R. B. Pinter and B. Nabet, Eds. CRC Press, Chapter 17, 451–462.

LAND, E. H. AND MCCANN, J. J. 1971. Lightness and retinex theory. *J. Opt. Soc. Am. 63*, 1, 1–11.

LEWIS, M. B. AND JOHNSTON, R. A. 1998. Understanding caricatures of faces. *Quart. J. Exper. Psych. 50A*, 2, 321–346.

MARR, D. 1982. *Vision, A Computational Investigation into the Human Representation and Processing of Visual Information*. W H Freeman and Company, San Fransisco, Calif.

MARR, D. AND HILDRETH, E. C. 1980. Theory of edge detection. *Proc. Roy. Soc. London, B 207*, 187–217.

MAURO, R. AND KUBOVY, M. 1992. Caricature and face recognition. *Mem. Cogni. 20*, 4, 433–440.

OSTROMOUKHOV, V. 1999. Digital facial engraving. *Proceedings of SIGGRAPH 99*. ACM, New York, 417–424.

PALMER, S. E. 1999. *Vision Science: Photons to Phenomenology*. The MIT Press, Cambridge, Mass.

PEARSON, D. E., HANNA, E., AND MARTINEZ, K. 1990. Computer-generated cartoons. In *Images and Understanding*. Cambridge University Press, Cambridge, Mass.

PEARSON, D. E. AND ROBINSON, J. A. 1985. Visual communication at very low data rates. *Proc. IEEE 73*, 4 (April), 795–812.

PESSOA, L., MINGOLLA, E., AND NEUMANN, H. 1995. A contrast- and luminance-driven multiscale network model of brightness perception. *Vis. Res. 35*, 15, 2201–2223.

RAMACHANDRAN, V. AND HIRSTEIN, W. 1999. The science of art a neurological theory of esthetic experience. *J. Conscious. Stud. 6*, 6–7, 15–51.

REDMAN, L. 1984. How to draw caricatures. In *Aspects of Face Processing*. Contemporary Books, Chicago Ill.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. In *ACM Transactions on Graphics, Proceedings SIGGRAPH 2002*. http://www.cs.ucf.edu/~reinhard/cdrom/, ACM, New York.

RHODES, G., BRENNAN, S., AND CAREY, S. 1987. Identification and ratings of caricatures: Implications for mental representations of faces. *Cogn. Psych. 19*, 473–497.

RHODES, G. AND TREMEWAN, T. 1994. Understanding face recognition: caricature effects, inversion, and the homogeneity problem. *Vis. Cogn. 1*, 2/3.

RHODES, G. AND TREMEWAN, T. 1996. Averageness, exaggeration, and facial attractiveness. *Psych. Sci. 7*, 2 (Mar.), 105–110.

SALISBURY, M., ANDERSON, C., LISCHINSKI, D., AND SALESIN, D. H. 1996. Scale-dependent reproduction of pen-and-ink illustrations. In *SIGGRAPH 96 Conference Proceedings*. ACM, New York, 461–468.

SALISBURY, M. P., ANDERSON, S. E., BARZEL, R., AND SALESIN, D. H. 1994. Interactive pen-and-ink illustration. In *Proceedings of SIGGRAPH 94*. ACM, New York, 101–108.

SALISBURY, M. P., WONG, M. T., HUGHES, J. F., AND SALESIN, D. H. 1997. Orientable textures for image-based pen-and-ink illustration. In *Proceedings of SIGGRAPH 97*, ACM, New York, 401–406.

SOUSA, M. C. AND BUCHANAN, J. W. 1999. Observational model of blenders and erasers in computer-generated pencil rendering. In *Graphics Interface '99*. ACM, New York, 157–166. ISBN 1-55860-632-7.

STEVENAGE, S. V. 1995. Can caricatures really produce distinctiveness effects? *Brit. J. Psych. 86*, 127–146.

SWELLER, J. 1972. A test between the selective attention and stimulus generalisation interpretations of the easy-to-hard effect. *Quart. J. Exper. Psych. 24*, 352–355.

TANAKA, T. AND OHNISHI, N. 1997. In Painting-like Image Emphasis based on Human Vision Systems. In *Proceedings of Eurographics '97 16*, 3 (Aug.), 253–260.

TVERSKY, B. AND BARATZ, D. 1985. Memory for faces: Are caricatures better than photographs? *Mem. Cognit. 13,* 45–49.

VALENTINE, T. 1991. A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quart. J. Exper. Psych. 43A*, 2, 161–204.

WANDELL, B. A. 1995. *Foundations of Vision*. Sinauer Associates, Inc.

WINKENBACH, G. AND SALESIN, D. H. 1994. Computer-generated pen-and-ink illustration. In *Proceedings of SIGGRAPH 94*. ACM, New York, 91–100.

WONG, E. 1999. Artistic rendering of portrait photographs. M.S. thesis, Cornell University.

YARMEY, A. D. 1973. I recognize your face but I can't remember your name: Further evidence on the tip-of-the-tongue phenomenon. *Mem. Cogn. 1*, 287–290.

YOUNG, A. W., HAY, D. C., AND ELLIS, A. W. 1985. The face that launched a thousand slips: Everyday difficulties and errors in recognizing people. *Brit. J. Psych. 76*, 495–523.

Received May 2002; revised December 2002; accepted February 2003