# Perception-Driven Black-and-White Drawings and Caricatures

*Bruce Gooch Erik Reinhard Amy Gooch*

UUCS-02-002

School of Computing
University of Utah
Salt Lake City, UT 84112 USA

January 22, 2002

## *Abstract*

In this paper, we present a method for automatically creating black-and-white drawings of human faces from photographs. We then demonstrate new techniques for deforming these drawings to create caricatures intended to highlight and exaggerate facial features. A number of psychophysical studies were performed that show that both the black-and-white drawings and the caricatures generated using our techniques have similar effectiveness as photographs in recognition tasks, and perform better than photos in learning tasks. Our methods are useful for a number of potential applications, ranging from forensics to telecommunications and from cartoons to education and psychology research.

# Perception-Driven Black-and-White Drawings and Caricatures

Bruce Gooch    Erik Reinhard    Amy Gooch

University of Utah

## Abstract

In this paper, we present a method for automatically creating black-and-white drawings of human faces from photographs. We then demonstrate new techniques for deforming these drawings to create caricatures intended to highlight and exaggerate facial features. A number of psychophysical studies were performed that show that both the black-and-white drawings and the caricatures generated using our techniques have similar effectiveness as photographs in recognition tasks, and perform better than photos in learning tasks. Our methods are useful for a number of potential applications, ranging from forensics to tele-communications and from cartoons to education and psychology research.

**CR Categories:** I.3.7 [Computing Methodologies]: Computer Graphics—3D Graphics; I.4.10 [Computing Methodologies]: Image Processing and Computer Vision—Image Representation

**Keywords:** Black-and-white drawings, Caricatures, Human visual perception, Non-photorealistic rendering

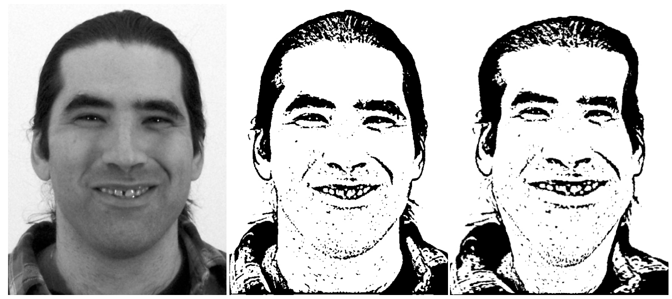

Figure 1: *Example of using (a) a photograph to derive (b) a black-and-white drawing and (c) a caricature.*

## 1 Introduction

Caricatures constitute a powerful medium to express and exaggerate distinctive features of human faces. They are usually created by skilled artists who use lines to represent facial features. We view caricatures to be drawings with extreme exaggerations of facial features. Previous caricature images are usually based on line drawings and have lines that are one pixel wide. Our caricatures are based on black-and-white drawings with varying line widths and parts of the drawing may be filled in.

It could be argued that line drawings as well as caricatures derived from such line drawings form impoverished environments when compared with their photographic counterparts. Within such impoverished environments, previous research has shown that caricatures can be recognized faster than line drawings that accurately portray someone's face (these are called veridical line drawings) [Brennan 1985; Rhodes et al. 1987; Benson and Perret 1991; Stevenage 1995]. This is known as the super portrait effect. Similarly, line-drawn caricatures tend to be learned faster in learning tasks than veridical line drawings [Stevenage 1995]. On the other hand, when caricatures are compared with photographs, these effects tend to be either minimal or completely absent [Benson and Perret 1991].

We postulate that the reason for this discrepancy results from the fact that line drawings have too little information content to be recognized as quickly as photographs. Hence, in this paper we seek to develop a method to first compute black-and-white drawings from photographs (Figure 1a and b) without removing too much information to have an adversary effect on recognition speed. Second, these drawings are then warped into caricatures (Figure 1c). Although our caricature generation algorithm requires a small amount of user intervention, our methods do not rely on skilled users to provide input, making them suitable for a wide variety of applications.

Both steps of our process to generate caricatures are based on aspects of human vision. The early stages of the human visual system reduce the amount of information present in the image [Atick and Redlich 1992], due to the fact that the number of photo-receptors in the retina is an order of magnitude larger than the number of nerves leaving the eye [Palmer 1999]. By processing photographs in a similar way, we can ensure that we only discard information that will not be processed by the human visual system. The result is an image that resembles a drawing with line widths that may vary and with some sections of the drawing filled in. The method is described in detail in Section 2.

The next step of processing warps the drawing to produce a caricature. It is argued that humans perform face recognition using average based coding, i.e., the distance of a given face in some feature space to an average face is encoded [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 account for the same effects observed in face recognition tasks, but in our opinion the average based coding paradigm lets itself be cast more easily into a computational model. To create a caricature, the difference between an average face and a particular face can be computed for various facial features and be exaggerated by a specified amount. We provide methods to automatically deviate from an average face, as well as techniques that allow meaningful free-form warping to allow more extreme caricaturing. This approach is outlined in Section 3.

The advantages of our approach are that the resulting images can be encoded in very little space, making them suitable for rapid transmission over low band-width networks. We foresee applications in telecommunications where users may transmit cartoon-like signatures of themselves when making phone calls using their PDA's. Second, since learning tasks can be sped up by using our drawings, this may benefit certain visual learning applications. For the same reason our methods may have applications in forensics, for example by constructing super portraits from photographs of missing persons. Finally, research in face recognition may benefit from using our techniques, rather than relying on single width line drawings as used in previous research.

To demonstrate the usefulness of our approach to producing black-and-white drawings and caricatures, we validated them using a set of psychophysical experiments. Although the images created with our techniques have a reduced information content, these experiments show that recognition tasks are largely unaffected, while learning speed is improved, presumably due to the fact that all necessary visual cues are preserved in our black-and-white drawings. The validation experiments are described in Section 4.

# 2 Black-and-white drawings

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; Salisbury et al. 1996; Salisbury et al. 1994; Winkenbach and Salesin 1994]. With this idea in mind, we would like to produce easily recognizable black-and-white drawings of faces. Some parts of the drawing may be filled in if this increases recognizability. In addition, we would like to be able to derive such drawings from photographs without skilled user input.

Computing a black-and-white drawing from a photograph can be done in many different ways. One possible method is to threshold luminance information, which captures the global arrangement of light and dark features [Pearson and Robinson 1985]. However, as demonstrated below, this approach fails to preserve important high luminance details, and thus only captures some of the facial features.

For machine vision applications, edge detection algorithms are often used. However, most edge detection algorithms produce thin lines that are connected. While this is a basic requirement in machine vision, it is specifically not needed for black-and-white drawings and may even reduce recognizability [Davies et al. 1978]. However, edge detection algorithms are often based on computational models of human vision. The processing that takes place in the early stages of human vision appears to produce imagery that resembles line drawings.

Our algorithm for creating drawings is therefore based on a model of human brightness perception to capture the fine detail of face images. It is augmented with a thresholded luminance image to fill in the global structure of the face. The general approach of our method is outlined in Figure 2. In the following sections we briefly introduce the concept of brightness perception, followed by the computational model of our drawing algorithm.

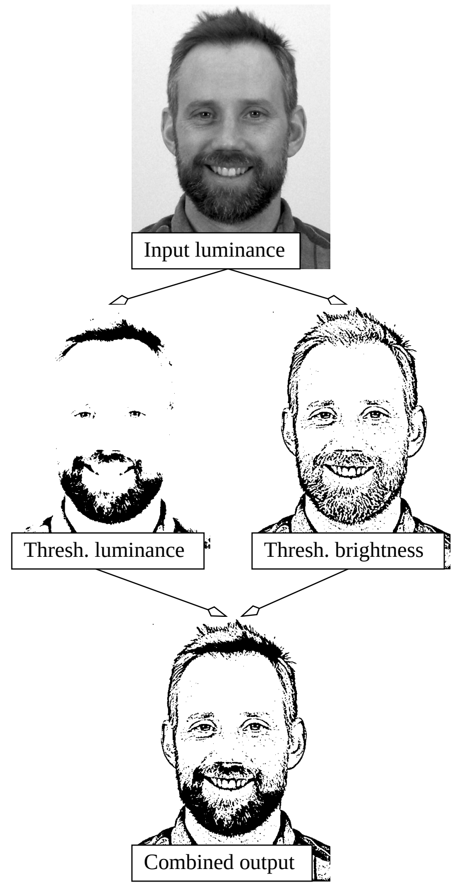


Figure 2: *Brightness is computed from a photograph, then thresholded and multiplied with thresholded luminance to create a black-and-white drawing.*

## 2.1 Contrast and brightness perception

While light is necessary to convey information from objects to the retina, the human visual system attempts to discard certain properties of light [Blommaert and Martens 1990; Atick and Redlich 1992]. An example is brightness constancy, where brightness is defined as a measure of how humans perceive luminances falling on the retina. It is therefore the subjective sensation of an object's albedo.

Brightness perception can be modeled using operators such as differentiation, integration and thresholding [Land and McCann 1971; Arend and Goldstein 1987]. These methods frequently model lateral inhibition which is one of the most pervasive structures in the visual nervous system [Palmer 1999]. Lateral inhibition is caused by a given cell's receptive field having a center-surround organization. Thus cells 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 [Cohen and Grossberg 1984; Blommaert and Martens 1990; Gove et al. 1995; Pessoa et al. 1995; Hansen et al. 2000], Gaussian smoothed Laplacians [Marr and Hildreth 1980; Marr 1982] and Gabor filters [Jernigan and McLean 1992].

Early and well-known computational models of brightness perception are the retinex theory [Land and McCann 1971], and Horns model for computing lightness [Horn 1974]. 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 an image and detects zero crossings in the result. Note that the second derivative can be closely approximated by computing the difference of two Gaussian blurred images if the Gaussians are scaled by a factor of $1.6$ with respect to each other [Marr 1982], a feature employed in our computational model.

## 2.2 Computing a black-and-white drawing

We adapt Blommaert and Martens' model of human brightness perception [Blommaert and Martens 1990] to create black-and-white drawings from photographs. This model is aimed at understanding the perception of brightness 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. To a first approximation, the receptive fields in the human visual system are isotropic with respect to brightness perception, and so they can be modeled by circularly symmetric Gaussian profiles $R_i$:

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

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 that is evoked at different 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 Equations 1 and 2. Subtracting $V_1$ and $V_2$ therefore 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. The parameter $\phi$ is typically set to 1, but can be varied to manipulate the amount of fine detail that will be present in the drawing. An expression for brightness $B$ is now derived by summing $V$ over all eight scales:

$$B(x, y) = \sum_{s=s_0}^{s_1} V(x, y, s), \tag{4}$$

with the lower and upper boundaries set to $s_0 = 1$ pixel and $s_1 = 1.6^8 = 43$ pixels. For computational convenience, the scales $s$ are spaced by a factor of $1.6$ with respect to each other. Finally, this brightness image can be thresholded by computing the average brightness of the image and setting all pixels that are above threshold to white and all other pixels to black. The resulting image can be interpreted as a black-and-white drawing, although we find that filling in the dark parts produces images that are more suitable as input to the caricaturing process. This filling in can be accomplished by thresholding the luminances of the input image separately and multiplying the result of this operation with the thresholded brightness image [Pearson and Robinson 1985]. The process of computing such a drawing is illustrated in Figure 2.

### 2.3 Results

Applying a model of human brightness perception leads to images that can be perceived as black-and-white drawings. Some of the results are shown in Figure 3. These drawings are based on photographs but contain much less information. For example, shading is completely removed from the image. As such, the storage space required for these drawings compares favorably with the space needed to store photographs. It is generally accepted that photographic material is most efficiently stored in lossy file formats such as jpegs. The drawings created with the method described in this section are most efficiently stored as tiffs, using CCITT group 4 compression [International Telecommunications Union 1988]. Table 1 shows the number of bits per pixel required to store photographs and drawings of the same size in their respective most optimal format. As this table indicates, the black-and-white drawing requires between six and twelve time less storage space than the grey-scale photograph it was derived from.

Finally, on a 400 MHz R12k processor, the computation time for a $1024^2$ drawing 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 Gaussian blurred images. 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 multi-resolution spline method may yield satisfactory results [Burt and Adelson 1983].

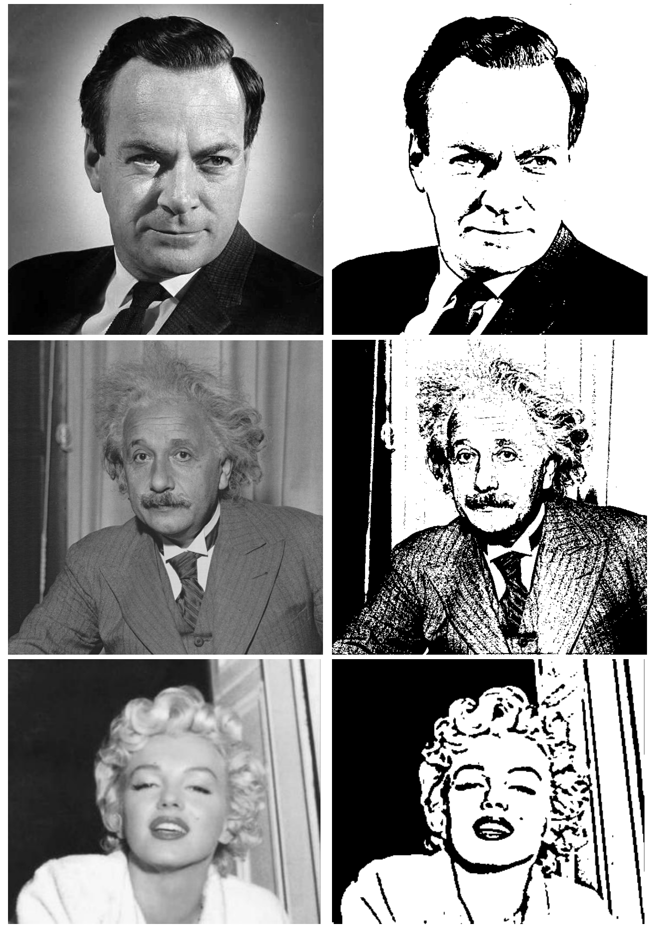

Figure 3: *Some results of our perception based drawing algorithm.*

| Size (pixels) | 429x619 | 320x240 | 160x160 |
|---|---|---|---|
| Photograph | 1.21 | 0.96 | 1.20 |
| Drawing | 0.10 | 0.11 | 0.19 |

Table 1: *Storage space (in bits per pixel) for photographs and drawings computed using our method*

## 3 Caricatures

Caricatures are traditionally the domain of skilled artists who devote much of their careers to learning which aspects of a human face to exaggerate. Automatically creating such drawings has thus far been an elusive goal, and attempts to automate aspects of this process are few and far between. The most well-known attempt is called the “Caricature Generator” [Brennan 1985], which is based on the notion of an average face. The positions of $165$ feature points, which are indicated by the user by clicking on a photograph, are compared with the positions of these points on an average face. By moving these points away from the average, an exaggeration can be created. A veridical line drawing is created by connecting the feature points with splines before translating them. A caricature is created by doing the same after translating the feature points over some distance. This method was later extended to allow the same warp to be applied to the input image to produce a photographic quality caricature [Benson and Perret 1991]. The Caricature Generator is used in many psychophysical experiments and has become a *de facto* standard for those conducting research in face recog-

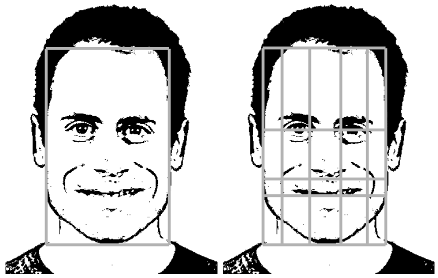

Figure 4: *Left: the face is framed by four border lines. Right: facial features and interior lines are matched*

nition [Rhodes et al. 1987; Benson and Perret 1994; Rhodes and Tremewan 1994; Stevenage 1995; Rhodes and Tremewan 1996] (see also Section 4).

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 and skilled user input, which limits their applicability for every-day use. We propose a semi-automatic method to produce caricatures requiring only untrained user input. Our algorithm could potentially be automated when better feature finding algorithms become available. In the following, we describe an algorithm to create super portraits, which is based on a model of how humans are thought to recognize faces. Then, we extend this method to allow more dramatic user guided caricatures.

### 3.1 Creating a super portrait

It has been shown that humans recognize faces based on the amount that facial features deviate from an average face [Valentine 1991; Tversky and Baratz 1985]. Thus, to produce a super portrait, features are distorted based on how much these features deviate from an average or norm face [Brennan 1985].

To construct a super portrait, we note that a norm face can be constructed in different ways, for example by averaging the positions of facial features for a large number of people. Our approach on the other hand, is based on the following metric [Redman 1984]. First, the face is framed with four lines, as shown in Figure 4. We then specify four vertical lines that mark the inner and outer corners of the eyes. Next, three horizontal lines mark the position of the eyes, the tip of the nose and the mouth. A norm face is characterized by the fact that the vertical lines are equidistant, while the horizontal eye, nose and mouth lines are set at $4/9$, $6/9$ and $7/9$ from the top of the frame [Redman 1984]. We call this set of horizontal and vertical lines a feature grid.

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

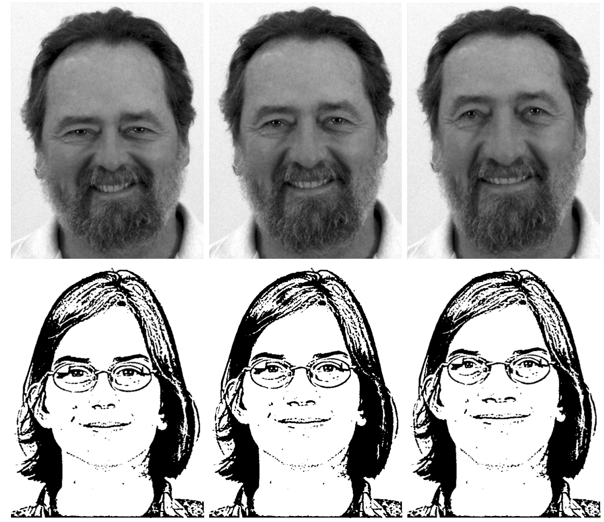

Figure 5: *The top row shows a 50% anti-caricature, the veridical photograph, and a 50% super portrait. The second example shows a 50% anti-caricature, the veridical line drawing, and a 50% super portrait.*

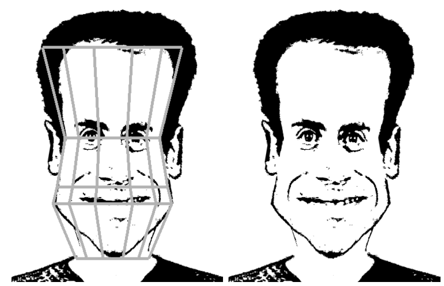

Figure 6: *Left: both grid and underlying image can be warped interactively. Right: the final image.*

### 3.2 Creating a caricature

For some applications, creating super portraits may not be sufficient and therefore we 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 6). 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 unskilled users from producing unrecognizable faces. In addition, the implementation proved straightforward, both in OpenGL and Java, and interactive manipulation was achieved on current workstations. Caricatures created by users who were given minimal verbal training (between 1 and 3 minutes) are presented in Figure 7.

## 4 Validation

The effectiveness of our drawings and caricatures is evaluated for specific tasks using psychophysical testing. In particular, we are interested in determining if recognition times of faces and the speed of learning new faces are affected by our black-and-white drawing and caricature algorithms. The hypothesis is that if our algorithm

Figure 7: *The top row shows the photographs used to create the caricatures in the remainder of the figure.*

does not affect the speed of recognition of familiar faces, then the information reduction achieved by our algorithm is benign and the resulting images can be used for tasks where recognition speed is essential.

Second, we ask the question if our drawing and caricature algorithms affect the speed 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]. This implies that learning to match names with unfamiliar black-and-white drawings of the type presented in this paper, as well as caricatures derived from them, may be easier than learning to match names from photographs of unfamiliar persons.

To test these hypotheses, we performed two experiments which are replications of earlier distinctiveness experiments [Stevenage 1995]. While these previous experiments assessed the effect of line drawings and line-drawn caricatures on recognition and learning speed, we use the same experiments to validate our black-and-white drawing and caricaturing techniques. In addition, we compare our drawings and caricatures with the photographs they are based on in terms of recognition and learning speeds.

Both experiments are described in detail in the appendices, while in the remainder of this section we give a broad overview of their set-up and present the results.

## 4.1 Recognition time

This experiment assesses the recognition time of familiar faces presented as photographs, drawings and caricatures. Based on the results obtained with the Caricature Generator, which employs line drawings and line-drawn caricatures [Brennan 1985; Rhodes et al. 1987; Benson and Perret 1991], it is expected that caricatures would elicit a faster response than black-and-white drawings. Images created by the Caricature Generator are made up of thin lines and such an impoverished medium may amplify this effect [Davies et al. 1978]. When recognition times for photographs are compared with those for photographic quality caricatures, then both photographs and caricatures tend to be recognized equally fast, provided that the caricature level remains below 48% [Benson and Perret 1991]. For a 48% caricature, the original photograph was recognized faster.

The above leads us to think that caricatures created using our technique may be recognized somewhat slower than both photographs and black-and-white drawings. To test this hypothesis, our first experiment is a replication of an earlier recognition time experiment [Stevenage 1995].

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 drawings, photographs and caricatures or drawings and caricatures. The details of this experiment are presented in Appendix A.

The difference between the mean recognition time for photographs and drawings of familiar faces was not statistically different, hovering around 1.84s. The caricatures were recognized around 5% slower than photographs. There was no statistical difference in the reaction time between caricatures and drawings. In each condition, the accuracy of recognition was higher than 98%, indicating that there is no speed for accuracy trade-off in this experiment.

Thus, we conclude that the strategy we have chosen to create drawings from photographs has no impact on recognition speed and can therefore be used for tasks in which speed of recognition is important. 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 in situations where a slightly increased reaction time can be tolerated and silliness is a positive attribute.

### 4.2 Learning speed and accuracy

While some research has found no advantage to using caricatures in a learning paradigm [Hagen and Perkins 1983], others have shown that caricatures of unfamiliar faces can be learned in fewer trials than the same faces shown as veridical images [Mauro and Kubovy 1992; Stevenage 1995].

In this experiment, each subject was presented with images of twelve unfamiliar faces in sequence, which were verbally assigned names. Each subject was shown either photographs, drawings, or caricatures, but not a combination of these. Next, the images were shown again, and the subject was asked to give the name for each image. During each iteration of this process, the subject was told which mistakes he or she had made. The images were then shuffled, and the process was repeated until the subject could correctly name all twelve faces. Compared with photographs, learning to recognize twelve unfamiliar faces was more than twice as fast for trials with black-and-white drawings. Caricatures were over 1.5 times faster than photographs. The details of this experiment are presented in Appendix B.

To test learning accuracy, subjects who were trained using the drawing or caricature images participated in a follow-up experiment using the original photographs in an otherwise identical setup. In this experiment, the number of incorrectly named faces was recorded. Both the training using caricatures and black-and-white drawings resulted in a 98% naming accuracy. Details are given in Appendix C.

The rate of learning can thus be increased by a significant amount using our algorithms without 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]. The information reduction achieved by our algorithms appears to have reached this goal.

## 5 Conclusions

Humans appear to employ dedicated processing for face recognition [Ellis 1986; Biederman and Kalocsai 1998; Gauthier et al. 1999], which makes compressing photographs of human subjects without losing important information a difficult task. The human visual system reduces the information content of the images that fall on the retina and this paper has shown that information reduction of photographs based on the human visual system is possible without impacting certain tasks, and even improves performance in other tasks. This leads to a compression algorithm resembling black-and-white drawings that allows a six to twelve-fold compression over the widely used jpeg compression on the associated photographs. In addition, our validation experiments have shown that there is not a measurable impact on face recognition speed, while learning tasks are significantly faster for the line drawings. We believe this to be due to the fact that only the necessary visual cues are retained in the drawings.

In addition, our methods do not require skilled user input from artists for either the drawing stage or the caricature stage. This makes the work presented in this paper suited for a wide variety of potential applications in fields ranging from telecommunication to education, forensics and psychology research.

## Appendix A: Recognition speed

*Subjects.* Forty two 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, a drawing, and as a caricature. The photographs were taken indoors using a Kodak 330 digital camera using its flash. In a pilot study five independent judges rated all drawings and caricatures as good likenesses of the faces they 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 subtended a visual angle of 12.9 degrees. Images were shown for five seconds at five second intervals.

*Procedure.* We conducted three two-part experiments, each with 14 subjects. The first part allowed subjects to rate their familiarity with a list of 20 names on a 7 point scale with a purpose designed user interface. Subjects 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 [Yarmey 1973; Young et al. 1985; Stevenage 1995].

In the second part of this experiment, the 12 highest rated faces are selected for each participant and were shown in two of three possible conditions. Subjects in Experiment A.1 saw photographs and drawings. Experiment A.2 consisted of photographs and caricatures, and Experiment A.3 consisted of drawings and caricatures. The written instructions for this part are: "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

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

Table 2: *Recognition speed results, showing the minimum, maximum, and mean time over average subject data for each condition in each experiment.*

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

Table 3: *Learning speed experiments, showing the minimum, maximum, and mean number of trial iterations for the experiment presented in Appendix B.*

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 drawings ($p = 0.55$) and a marginal advantage for naming drawings compared to caricatures ($p = 0.07$). The accuracy for recognizing photos, drawings and caricatures are 98%, 99% and 98% respectively. Table 2 provides minimum, maximum, and mean times recorded for each condition on each experiment.

## Appendix B: Learning speed

*Subjects.* Thirty 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 drawings and 12 caricatures derived from these photos were all rated as good likenesses. All photos, drawings and caricatures were printed on a laser printer at a size of 6”x8” 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 found on the web [Cherished Moments]. 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 three groups of 10 and each participant was individually presented with either photographs, drawings 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. 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 3. Drawings are learned significantly faster than photos ($p < 0.001$). Caricatures versus photos and drawings versus caricatures could not be distinguished statistically ($p = 0.093$, $p = 0.081$). We conclude that learning was quickest when the faces were presented as black-and-white drawings, followed by caricatures and then photographs.

## Appendix C: Learning Accuracy

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

*Materials.* The same materials were used as in Experiment B.

*Procedure.* In this experiment we explored whether caricatures and drawings 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.* The number of incorrectly identified faces was recored for each subject given either the drawing or caricatured training. In both cases the accuracy was 98%. Hence, there was no difference in accuracy between subjects trained with drawings or caricatures. We this conclude that the transfer of learning to photographs appears to be equally good and above chance.

## 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. *Perception and Psychophysics 43*, 65–80.

ATICK, J. J., AND REDLICH, N. A. 1992. What does the retina know about natural scenes? *Neural Computation 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 psuchological analysis of object and face recognition. In *Face Recognition: From Theory to Applications*, Springer-Verlag, 3–25.

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

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 Transactions on Graphics 2*, 4, 217–236.

CHERISHED. Moments. http://www.cherishedmoments.com/most-popular-baby-names.htm.

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

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

ELLIS, H. D. 1986. Introduction to aspects of face processing: ten questions in need of answers. In *Aspects of Face Processing*, Nijhoff, Dordrecht, H. Ellis, M. Jeeves, F. Newcombe, and A. Young, Eds., 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. *Nature Neuroscience 2*, 6 (June), 568–573.

GOVE, A., GROSSBERG, S., AND MINGOLLA, E. 1995. Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience 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, T., BARATOFF, G., AND NEUMANN, H. 2000. A simple cell model with dominating opponent inhibition for robust contrast detection. *Kognitionswissenschaft 9*, 93–100.

HORN, B. K. P. 1974. Determining lightness from an image. *CVGIP 3*, 277–299.

INTERNATIONAL TELECOMMUNICATIONS UNION, 1988. Facsimile coding schemes and coding control functions for group 4 facsimile apparatus. Recommendation T.6.

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, ch. 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. *The Quarterly Journal of Experimental Psychology 50A*, 2, 321–346.

MARR, D., AND HILDRETH, E. C. 1980. Theory of edge detection. *Proceedings of the Royal Society of London, B 207*, 187–217.

MARR, D. 1982. *Vision, a computational investigation into the human representation and processing of visual information*. W H Freeman and Company, San Fransisco.

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

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

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

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. *Vision Research 35*, 15, 2201–2223.

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

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

RHODES, G., AND TREMEWAN, T. 1996. Averageness, exaggeration, and facial attractiveness. *Psychological Science 7*, 2 (March), 105–110.

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

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

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

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

STEVENAGE, S. V. 1995. Can caricatures really produce distinctiveness effects? *British Journal of Psychology 86*, 127–146.

SWELLER, J. 1972. A test between the selective attention and stimulus generalisation interpretations of the easy-to-hard effect. *Quarterly Journal of Experimental Psychology 24*, 352–355.

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

VALENTINE, T. 1991. A unified account of the effects of distinctiveness, inversion and race in face recognition. *The Quarterly Journal of Experimental Psychology 43A*, 2, 161–204.

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

YARMEY, A. D. 1973. I recognize your face but I can't remember your name: Further evidence on the tip-of-the-tongue phenomenon. *Memory and Cognition 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. *British Journal of Psychology 76*, 495–523.