HUMAN FACIAL ILLUSTRATIONS: CREATION AND EVALUATION USING BEHAVIORAL STUDIES AND FUNCTIONAL MAGNETIC RESONANCE IMAGING

by

Bruce Gooch

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Computer Science

School of Computing

The University of Utah

July 2003

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

SUPERVISORY COMMITTEE APPROVAL

of a dissertation submitted by

Bruce Gooch

This dissertation has been read by each member of the following supervisory committee and by majority vote has been found to be satisfactory.

<div></div>

_____	_____
	Chair: Richard F. Riesenfeld
_____	_____
	Peter Shirley
_____	_____
	Elaine Cohen
_____	_____
	Greg Jones
_____	_____
	Adam Finkelsein
_____	_____
	Sarah Creem-Regehr

FINAL READING APPROVAL

To the Graduate Council of the University of Utah:

I have read the dissertation of _____Bruce Gooch_____ in its final form and have found that (1) its format, citations, and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the Supervisory Committee and is ready for submission to The Graduate School.

_____   _____

Date   Richard F. Riesenfeld
                                                  Chair: Supervisory Committee

Approved for the Major Department

_____

Christopher R. Johnson
Chair/Director

Approved for the Graduate Council

_____

David Chapman
Dean of The Graduate School

ABSTRACT

This dissertation presents: a method for creating black-and-white illustrations and caricatures of human faces from source photographs; and series of perceptual studies aimed at evaluating the effectiveness of the resulting images relative to photographs. The illustrations are generated by superimposing two images: a thresholded image of the output of a computational brightness model, and a thresholded luminance image. In addition, a new interactive technique is demonstrated for deforming images of faces to create caricatures that highlight and exaggerate representative facial features. The photographs and black-and-white illustrations are evaluated via psychophysical studies to assess speed and accuracy in learning and recognition tasks. These studies show that the facial illustrations and caricatures generated using these techniques are as effective as photographs in the recognition tasks. In the learning studies, tasks involving illustrations or caricatures were performed significantly faster than the same tasks were performed with photographs. The recognition invariance effect is used as an experimental probe in a functional magnetic resonance imaging (fMRI) experiment. The results of this experiment indicate that viewers may process illustrations differently from photographs.

To Amy,

More than the Sun, the Moon, and all the Stars.

CONTENTS

CHAPTERS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

fMRI study. Jim Lee completed the data analysis for the normalized fMRI data, which can be an arcane task. Many thanks go out to the three of you for your help in this work.

Erik Reinhard and Amy Gooch are co-authors on a paper titled "Human Facial Illustrations: Creation and Evaluation." I would like to thank both of my paper co-authors for their help in what proved to be the early stages of this work.

CHAPTER 1

INTRODUCTION

The intent of both Non-Photorealistic Rendering (NPR) and Photorealistic Rendering is to visually communicate information. The Photorealistic Rendering community creates images by simulating the physics of light in mathematically modeled scenes. The goal of Photorealistic Rendering is to create images that are indistinguishable from photographs of real world scenes. In contrast, NPR is a newer and in many ways broader field whose community possess an assortment of image creation goals. NPR images are created using a variety of methods from the simulation of traditional artistic media to completely *ad hoc*. The only unifying principle in NPR is that all NPR images are created in order to stimulate the human visual system.

One of the most important skills for an artist is to learn is choosing the correct medium for a given subject. Artists make these choices guided by considerations such as aesthetic appeal and the effectiveness of the medium in communicating the required visual message. The directors of feature films integrate NPR methods with traditional film media to produce effects never before possible. Nearly every 3D rendering package now has a cartoon shading plug-in that empowers non-artists with the ability to create images in ways only professional artists were once able. NPR gives the graphics community freedom to choose media in addition to a camera for creating images.

A common assumption in the graphics community is that NPR involves simulating natural artistic media. This assumption is not surprising, because the first research in the field focused on reproducing traditional art forms, such as pen and ink, watercolor, and oil on canvas. Technology development in any field first seeks

Figure 1.1. Images demonstrating the difference between photographs and Non-Photorealistically rendered computer graphics images.

to imitate the previous mode of working. Thus, NPR seems to be following the usual development scheme. NPR currently embraces a wider scope of research. For example, a recent trend in NPR research is interactive NPR techniques. Current work focuses on the detection and rendering of feature lines to communicate shape. Silhouettes, surface and texture boundary lines, as well as creases are important for communicating the shape of an object.

The goal of generating computer images that are indistinguishable from photographs is essential for a host of applications including design, marketing, and the entertainment industry. In many applications, however, a non-photorealistic image has advantages over a photorealistic image. NPR images omit extraneous detail, focus attention on relevant features, clarify, simplify, and disambiguate shape, and show hidden parts. An example of a photograph and an architectural rendering are shown in Figure 1.1.

The control of detail in an image for purposes of communication is becoming the hallmark of NPR images. Often this control of image detail is combined with stylization to evoke the perception of complexity in an image without an explicit representation. NPR images also provide a more natural vehicle for conveying

information at different levels of abstraction and detail. Some additional occasions when a NPR image has an advantage are listed below.

1. Image Reproducibility – In a technical journal printed in black and white, fully shaded 3D geometry may not print well. For example photographic images do not copy or fax as well as line art images.

2. Medical Visualization – Researchers are focusing on providing NPR algorithms, which can be manipulated interactively, for real time visualizations of volume data. A good example is the visualization of electric fields inside the human body.

3. Communication of Abstract Ideas – The human visual system expects realistically rendered characters to behave realistically. Therefore, non-photorealistic animation can be used to express ideas beyond the physical and logical norm, in a way that is acceptable to a general audience. An example of this is force diagrams used in physics textbooks.

4. Evoking the Imagination – Simple line drawings can communicate abstract ideas in ways that a photograph cannot. In a photorealistic image, everything in the scene is rendered in fine detail, leaving little to the imagination. In comparison, by not depicting every detail, a non-photorealistic image allows the viewer to share in the interpretive process.

5. Animation – When creating an animation it is necessary to focus the attention of the audience on the relevant actions and elements in the scene. A viewer inspecting the fine details of a photorealistic scene can miss the big picture. Most non-photorealistic techniques employ an economy of line, limiting the detail in a scene, which makes directing the attention of the viewers easier for an animator.

6. Compression – By not depicting all the detail required for photorealistic images, non-photorealistically rendered computer graphics images typically

take less time to create, can be rendered to the screen faster, and use less storage space. For example, half-tone images yield the same shape from shading cues as traditionally rendered computer graphics images when viewed from a distance. However, the half-tone images require between one tenth and one one-hundredth of the storage space [4, 38, 79].

7. Communication of Design or Process Completeness – Photorealistic rendering implies an exactness and perfection that may overstate the fidelity of the simulated scene to a photograph. NPR can aid a viewer in understanding that the image they see is only an approximate depiction of a scene. An excellent example of this phenomena is architectural rendering. Architects have found that on-site building conditions and variations in regional building codes can lead to last minute changes in building plans. If clients are shown realistic images of the proposed building these last minute changes can come as a shock, leading to angry, disappointed clients. However, if the clients are shown Non-Photorealistic images of the proposed building clients tend to accept the design process as incomplete and the plans as changeable. Therefore, the clients usually accept on-site changes.

The early adoption and subsequent interest in Photorealistic Rendering by the graphics community is most likely due to the "mission statement" of Photorealistic Rendering: "Create an image that is indistinguishable from a photograph." This mission statement gives Photorealistic Rendering a visual "Turing test", and an easily defined metric for a successful image. Non-Photorealistic Rendering does not have a single mission statement. Instead, researchers are pursuing a number of image creation goals. The goals of NPR include simulating traditional artistic media, understanding the human visual system, communicating effectively with low bandwidth, abstracting images, enhancing learning, and improving user interaction.

In the computer graphics community, rendering is the process by which a virtual scene is converted into an image. Photorealistic rendering has come to mean images' depicting what is real based on physical simulations. While physical phenomenon

are objective and relatively predictable, photorealism is inherently costly due to the vast quantity of data that must be specified and processed. The main difficulty in striving for photorealism is that the physical world is complex. To render a computer-modeled scene realistically, the scene must approximate its real-world counterpart to a high degree of accuracy, resulting in extremely complex geometry. Therefore, the realistic depiction of complex objects always involves a series of trade offs based on the image creator's time and computational budget.

Images created using a photorealistic rendering process can be objectively measured or visually compared to a photograph of the physical process being simulated. In NPR the goal of the computer generated images is believability. A believable image is one that effectively communicates the intent of the image's creator. Believability is simpler to achieve than photorealism, because a believable model needs only to include the details representative of the intent behind the image.

NPR is concerned with images that are abstract, guided by processes employed by artists. In contrast to photorealism, in which the driving force is the modeling of physical processes and behavior of light, the processes of human perception drive NPR techniques. Following this trend, this dissertation presents two techniques that may be applied in sequence to transform a photograph of a human face into a caricature. First, extraneous information is removed 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 called 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 these algorithms are shown in Figure 1.2. Compared to previous caricature generation algorithms this method requires only a small amount of untrained user intervention.

Figure 1.2. Left: The source photographs. Center: Examples of illustrations generated using the software reported in this dissertation. Right: Caricatures created by exaggerating representative facial features. The photographs are part of the Aleix Face Database and are included here courtesy of Aleix Martinez of Purdue University. The entire face database is accessible online [1].

This dissertation presents four results:

1. An automated method to produce easily recognizable black-and-white illustrations of human faces from photographs.

2. A user-assisted method to produce caricatures from images of human faces.

3. The results of behavioral studies that show that in some tasks illustrations yield better user performance under an objective standard.

4. The results of a functional magnetic resonance imaging (fMRI) study indicating that illustrations and photographs may be processed differently when viewed by humans.

The first two results are presented in Chapter 2, and the last two results are presented in Chapter 3.

CHAPTER 2

COMPUTER-GENERATED FACIAL ILLUSTRATIONS

This chapter documents a method for computing black-and-white illustrations from source photographs and an interactive technique for creating caricatures from facial images. The chapter details how these methods are related to previous work on facial images in the fields of Computer Graphics, Psychology and Art.

2.1 Background

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 [105, 123–125, 130, 143]. This idea provides motivation for algorithms that 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 should be removed because they are not an intrinsic aspect of the face. Ideally, such illustrations could be computed directly 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. Such methods rely heavily on user input [39, 105, 127, 144]. Stroke-based methods are mainly concerned with determining stroke placement to maintain tonal values across the surface of an object. For the current application the method should be automated and do away with gray-scale information.

Another method for creating facial illustrations is to only draw pixels in the image with a high intensity gradient [109, 110]. 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, "holes" are sometimes created in areas that should be all dark. Filling in the dark regions seems to produce images that are more suitable. This filling in can be accomplished by thresholding the input luminance separately and multiplying the result of this operation with the thresholded brightness image [110].

A third approach is to use 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. In fact, connectedness may reduce the recognizability of facial images [37].

To comply with the requirements that the method needs minimal trained user input and produces easily recognizable images, the algorithm presented here is based on a computational 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, a sense of absolute luminance levels is preserved by thresholding the input image and adding this to the result. The general approach of the method is outlined in Figure 2.1.

Because humans appear to employ dedicated processing for face recognition [17, 42, 53], 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. The method reported in this work for producing illustrations from photographs achieves this by applying a modified model of human brightness perception to photographs. It should be noted that there are many reasonable brightness models available and that this is still an active area of research [20, 54, 107, 140].
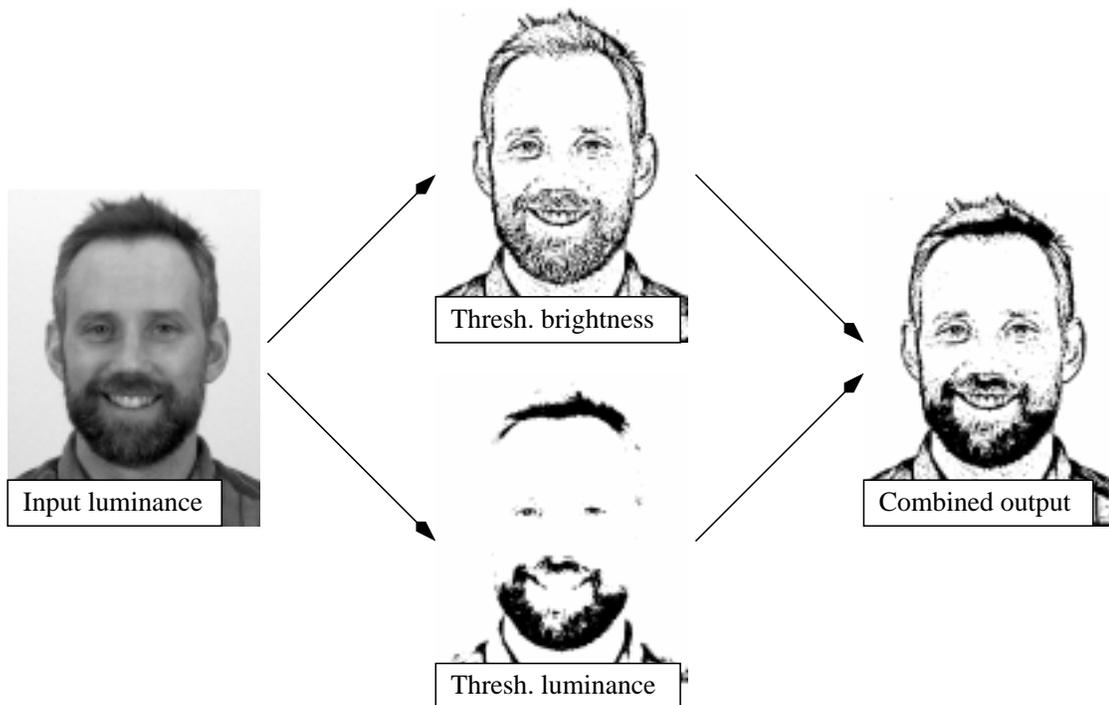
Figure 2.1. Brightness is computed from a photograph, then thresholded and multiplied with thresholded luminance to create an illustration.

2.2   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 [9, 19]. An example is brightness constancy, where brightness is defined as a measure of how humans perceive luminance [107].

Brightness perception can be modeled using operators such as differentiation, integration and thresholding [7, 83]. These methods model lateral inhibition, which is one of the most pervasive structures in the visual nervous system [107]. 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 that is bright in the center of the cells 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 [19, 33, 55, 63, 111], Gaussian smoothed Laplacians [92, 93] and Gabor filters [74].

Difference of Gaussians is a gray-scale image enhancement algorithm that involves the subtraction of one blurred version of an original gray-scale image from another, less blurred version of the original. The blurred images are obtained by convolving the original gray-scale image with Gaussian kernels having differing standard deviations. Blurring an image using a Gaussian kernel suppresses only high-frequency spatial information. Subtracting one image from the other preserves spatial information that lies between the ranges of frequencies that are preserved in the two blurred images. Thus, the difference of Gaussians is equivalent to a band-pass filter that discards all but a handful of spatial frequencies that are present in the original gray-scale image. The difference of Gaussians algorithm is useful for enhancing edges in noisy digital images because it removes high-frequency spatial detail including random noise. In its operation, the difference of Gaussians algorithm is believed to mimic how neural processing in the retina of the eye extracts details from images destined for transmission to the brain [19, 33, 55, 63, 111].

Gaussian smoothed Laplacians or a Laplacian-pyramid combine the advantages of predictive and transform methods of image processing [2, 25]. The Laplacian-pyramid is a versatile data structure with many attractive features for image processing. It has many applications such as progressive image transmission. The Gaussian smoothed Laplacian algorithm is implemented by first performing a low-pass filter (Gaussian filter) on the original image to get a reduced version of the image, this process is recursively repeated yielding a series of increasingly reduced images, that together are called a Gaussian pyramid. The Laplacian pyramid is a series of error images that are the difference between two levels of the Gaussian pyramid. The Laplacian pyramid is like a complement of the Gaussian pyramid; it can be decoded and recovered to the original image by expanding, then summing all the levels of the pyramid. The proper choice of the number and distribution of quantization levels must be carefully chosen to make the degradation imperceptible to a viewer [34, 56].

Gabor filters are bandpass filters that are used in image processing for feature extraction, texture analysis, and stereo disparity estimation. Multiplying a Gaussian function with a complex oscillation creates the impulse response of these filters. Gabor showed that these elementary functions minimize the space-time uncertainty product [51]. By extending these functions to two dimensions, it is possible to create filters that are selective for orientation [36]. Under certain conditions, the phase of the response of Gabor filters is approximately linear. This property is exploited by stereo approaches that use the phase-difference of the left and right filter responses to estimate the disparity in the stereo images. It has been shown that the profile of simple-cell receptive fields in the mammalian cortex can be modeled by oriented two-dimensional Gabor functions [6, 18, 80, 91, 112, 113, 122].

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 Hildreths zero-crossing algorithm [93]. 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. Therefore, LoG can 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 [92]; a feature also employed in the current computational model.

2.3   Computing Illustrations

To create illustrations from photographs, Blommaert and Martens [19] model of human brightness perception is adapted to an illustration algorithm. The aim of the Blommaert model is to understand brightness perception in terms of cell properties and neural structures. The scale invariance property of the human visual system can be modeled by assuming the outside world is interpreted at different levels of resolution. Blommaert and Martens [19] 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:

$$R_i(x, y, s) = \frac{1}{\pi(\alpha_i s)^2} \exp\left(-\frac{x^2 + y^2}{(\alpha_i s)^2}\right). \tag{2.1}$$

These Gaussian profiles operate at different scales s and at different image positions (x, y). R_1 is used 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 $\alpha_2 = 1.6\alpha_1$ is specified. A neural response V_i as function of image location, scale and luminance distribution L can be computed by convolution:

$$V_i(x, y, s) = L(x, y) \otimes R_i(x, y, s). \tag{2.2}$$

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

$$V(x, y, s) = \frac{V_1(x, y, s) - V_2(x, y, s)}{2^\phi/s^2 + V_1(x, y, s)}, \tag{2.3}$$

where center V_1 and surround V_2 responses are derived from Equations 2.1 and 2.2. Subtracting V_1 and V_2 leads to a Mexican hat shape that is normalized by V_1. The term $2^\phi/s^2$ is introduced to avoid the singularity that occurs when V_1 approaches 0 and models the (scale-dependent) rest activity associated with the center of the receptive field [19]. 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 this application deals with low dynamic range images, as well as an uncontrolled display environment (see below), the model is heuristically adapted by setting $\phi = 1$. In the course of the current work it was found that this parameter could 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:

$$B(x, y) = \sum_{s=s_0}^{s_{max}} V(x, y, s). \tag{2.4}$$

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 cannot be accurately controlled. For these reasons, the angles are translated 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 reuse of the surround computation at scale s_i for the center at scale s_{i+1}.

The result of these computations is an image that 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. In addition, areas with a constant gradient are removed becoming 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 is typically not shown in illustrations. Brightness images are typically gray with lighter and darker areas near regions with non-zero 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 the brightness representation into a two-tone image that resembles an illustration. This is accomplished 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. Average brightness is a good initial estimate in practice. All other constants in the brightness model are fixed as indicated above and therefore do not require further human intervention.

Figure 2.2 shows the result of thresholding the brightness representation and compares it to thresholding a high pass filtered image (middle) and an edge detected
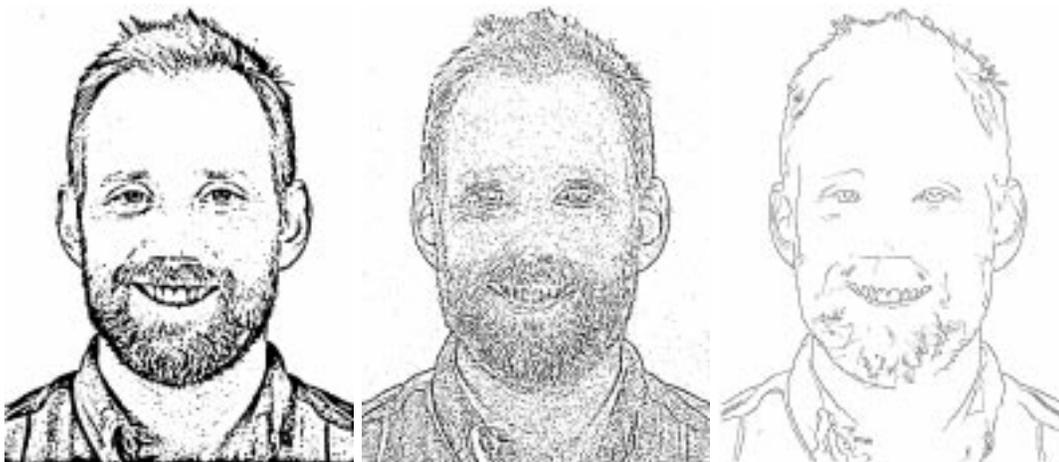
Figure 2.2. Left: Thresholded brightness. Center: High-pass filtered image followed by thresholding. Right: The output of Canny's edge detector [30].

image (right). Thresholding a high-pass filtered image and edge detecting are perhaps more obvious strategies that could potentially lead to similar illustrations. Figure 2.2 shows that this is not necessarily the case. The high-pass filtered image was obtained by applying a 5x5 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 [30] 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 filling in the dark parts produces images that seem to be 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 [110]. The threshold value is chosen manually according to taste, often falling in the range from 3% to 5% gray. The process of computing a portrait is summarized in Figure 2.1.

These facial illustrations are based on photographs but contain much less information, as shown in Figure 2.3. For example, shading is removed from the image, which is a property of Difference of Gaussians approaches. As such, the storage
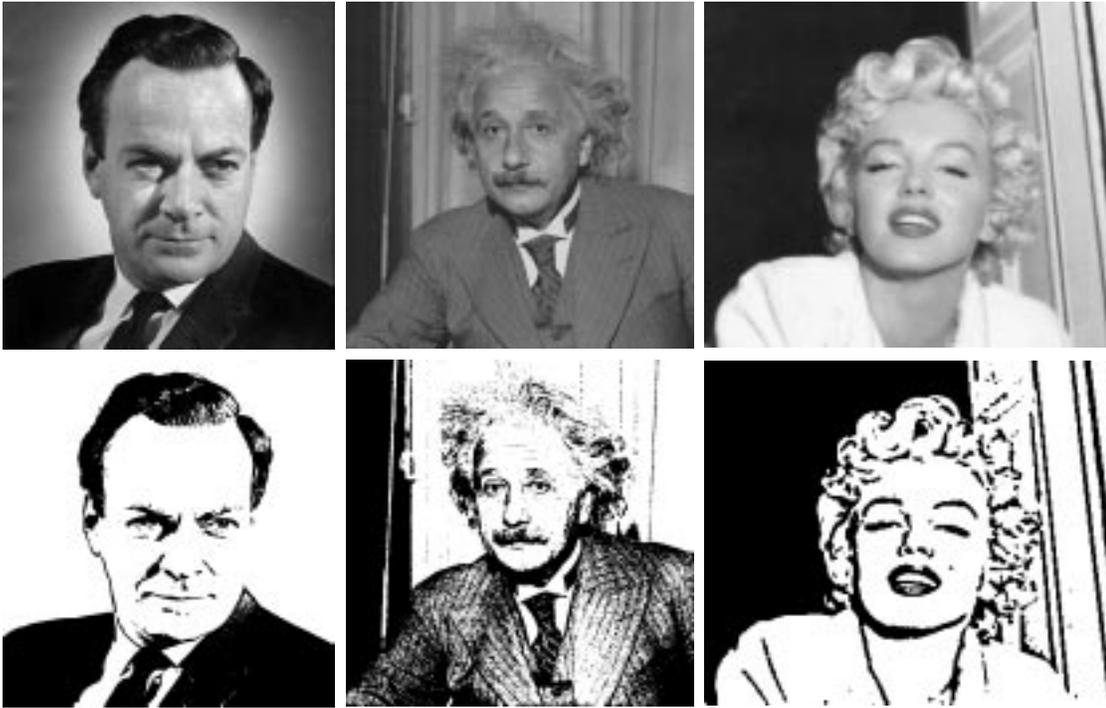
Figure 2.3. Top Row: Source images. Bottom Row: Results of the perception based portrait algorithm.

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

Table 2.1. This table shows the required storage space, in bits per pixel, for various sized photographs and facial illustrations.

space required for these illustrations is decreased from the space needed to store photographs.

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 Equation 2.2. Storage space requirements for photographs and facial illustrations are shown in Table 2.1.

The images could be computed faster with approximate methods, although this could be at the cost of some quality. In particular, a multi-resolution spline method may yield satisfactory results [26].

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. In addition, 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 usually improves the result. Further research may automate this process and so make this method better suitable for producing animated illustrations from video sequences. Better models of brightness perception, should they become available, may improve the 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 these results. Although Blommaert and Martens discuss a notion of absolute light levels [19], for the current 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.

While Figure 2.3 allows the reader to subjectively assess the performance of the algorithm, its real merit lies in the fact that specific tasks can be performed quicker using facial illustration images than when using photographs. This finding is presented in Chapter 3.

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

2.4   Superportraits and Caricatures

Two paradigms exist that explain how humans perform face recognition. In the average-based coding theory, the brain encodes a "feature space distance" from an average face to a given face [139]. An alternative model of face recognition is based on exemplars [87], 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.

2.4.1   Superportraits

Superportraits of human faces are a well-studied example of the peak shift effect in human visual perception [16, 117–119, 128]. The *peak shift* effect is a well-known principle in animal learning [62, 114]. 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. In the case of super portraits and caricatures, facial features that deviate from a population average are exaggerated, making the faces more easily recognized and learned. Figure 2.4 shows examples of superportraits.

2.4.2   Caricatures

The documented ability of caricatures to augment the communication content of images of human faces motivated the investigation of computer-generated caricatures [15, 22, 117, 128]. To create a caricature the difference between an average face and a particular face can be computed for various facial features. Traditionally skilled artists who use lines to represent facial features created caricatures. The

Figure 2.4. Left: Photographic examples. Right: Facial illustration examples. In both the photographic and facial illustration examples, the first images are of 50% anti-caricatures, the second images are the source images, and the third images are 50% superportraits.

skill of such artists 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 [22, 115, 117]. Automatically creating such drawings has been an elusive goal, and attempts to automate this process are sparse. Methods are described for automatically deviating from an average face, as well as techniques that allow meaningful warping to perform extreme caricaturing.

It has been postulated that humans recognize faces based on the amount that facial features deviate from an average face [134, 139]. Thus, to produce a caricature, features are exaggerated based on how far the face's features deviate from an average or norm face [22]. The most well known attempt is the "Caricature Generator" based on the notion of an average face [22]. Examples of facial illustrations and caricatures created using an implementation of the "Caricature Generator" software are shown in Figure 2.5. 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

Figure 2.5. Left: The source photograph. Center: An example of a line art image created using an implementation of the "Caricature Generator" software. Right: An example of a caricature created using an implementation of the "Caricature Generator" software.

to produce a photographic caricature [15]. The "Caricature Generator" has become a *de facto* standard for conducting research in face recognition [16, 117–119, 128].

A second semi-automated caricature generator is based on simplicial complexes [3]. 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 warps that are more general. This system is capable of interactively producing extreme exaggerations of facial features, but requires an experienced user 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. The methods in this work are semi-automatic and produce both superportraits as well as caricatures with less user training.

2.5   Creating a Superportrait

Superportraits are created in a semi-automatic way. A face is first framed with four lines, which comprise a rectangle, as shown in Figure 2.6. Four vertical lines are then introduced marking the inner and outer corners of the eyes, respectively.
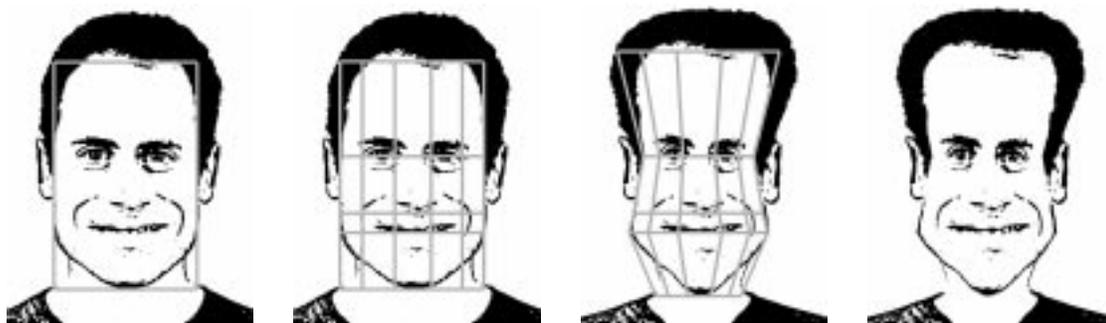
Figure 2.6. First: the face is framed by four borderlines. Second: facial features and interior lines are matched. Third: both grid and underlying image are warped interactively. Fourth: the resulting caricature.

Next, three additional horizontal lines mark the position of the eyes, the tip of the nose, and the mouth. This set of horizontal and vertical lines is called a facial feature grid (FFG).

To generate a FFG for a norm face a previously defined metric [115] is applied. 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 [115], but for the purpose of this work, the input is constrained 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 FFG can be exaggerated by computing the vectors between corresponding vertexes 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 superportrait, whereas negative percentages give rise to anti-caricatures, images that are closer to the norm face than the input image (Figure 2.4).

2.6   Creating a Caricature

As a manifestation of the peak shift effect, superportraits are useful in the study of human perception. However, more extreme distortions may be required in other applications. Therefore, the algorithm is extended to allow more freedom. Based

Figure 2.7. The exaggeration of facial features used by a professional caricature artist can easily be obtained by a novice user with the algorithm. First and third images: Examples of caricatures by noted caricature artist Ric Machin. Second and fourth images: Examples of caricatures created using the interactive software presented in this dissertation.

on the feature grid as described previously, vertices on both the right and left edges of the grid may be manipulated interactively. In addition, grid lines may be moved. These user actions are interactively visualized by warping the image according to the position of the vertexes (Figure 2.6). This process is constrained by disallowing the selection and manipulation of internal vertices. The resulting system is flexible enough to create amusing caricatures, while at the same time protecting untrained users from producing unrecognizable faces. Figure 2.7 compares examples of caricatures drawn by noted caricature artist Ric Machin to images created using this system.

These examples demonstrate that warping faces using a feature grid produces exaggerated images similar to those created by a professional artist. The implementation is straightforward, in both OpenGL and Java, and interactive manipulation was achieved on current workstations. 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. Caricatures created by users who were given minimal verbal training are presented in Figure 2.8.

Figure 2.8. 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.

2.7   Summary

This chapter presented a method for computing black-and-white illustrations from source photographs and an interactive technique for creating caricatures from facial images. While the illustrations and caricatures presented in this chapter may be subjectively good. In the next Chapter, the illustrations and caricatures are objectively evaluated in behavioral studies involving learning and recognition tasks.

CHAPTER 3

EVALUATION

Many studies have shown that learning in simplified environments may proceed faster than for similar learning tasks executed in the full environment [22, 117]. As llustrations and caricatures may be regarded as simplified with respect to photographs, learning tasks using them may be easier than the same task using the associated photographs. This chapter presents the results of perceptual studies and an fMRI study carried out using illustrations and caricatures as stimulus.

3.1 Background

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 behavioral study is conducted whereby participants perform specific tasks on sets of visual stimuli. Relative task performance is then related to the effectiveness of the image. If participants are statistically better at performing such tasks given a certain type of image, then these image types are said to be better at communicating for the given task.

Three studies have documented ability of caricatures to augment the communication content of images of human faces [15, 22, 117]. In these studies Caricature Generator software was used to produce images from source photographs. The Caricature Generator images were then systematically warped to exaggerate the difference of an individual face from a normalized face.

The accuracy of study participants in recognizing warped and unwarped images was compared. One study compared the accuracy of study participants in recognizing facial images drawn by a professional artist to single line weight images traced from the source photographs [128].

The learning speedup and the recognition invariance demonstrated in the behavioral studies suggest that different brain structures may be involved in the perception of artistic images than are involved in the perception of photographs. Functional magnetic resonance imaging (fMRI) is a method of determining brain activity based on a given stimulus. In this dissertation fMRI is treated as a *black box*, which will create images showing statistically valid regions of brain activation.

During an (fMRI) study, the brain of the participant is scanned repeatedly, usually using the fast imaging technique of echo planar imaging (EPI). The participant is required to carry out some task consisting of periods of activity and periods of rest. During the activity, the magnetic resonance (MR) signal from the region of the brain involved in the task normally increases due to the flow of oxygenated blood into that region. Simply put, an increase in neuronal activity is associated with an increase in local blood flow, and this increase is measured with fMRI. Signal processing is then used to reveal these regions. An fMRI study was undertaken to find if the illustrations stimulated different brain activity than the source photographs.

3.1.1   Statistical Significance

A common tool in modern research is statistical hypothesis testing. This process involves making a hypothesis and then collecting data to test that hypothesis. The method takes as its origin a re-formulation of the original research hypothesis, called a *null hypothesis*, which is commonly written H0. In its simplest version, a *null hypothesis* implies that the groups to be compared by means of the experiment are similar. In statistical terms, this is expressed by saying that the two data sets to be compared could very well have been drawn at random from the same set of data.

A P-value is a measure, assuming H0 is true, of how likely a given result is. A P-value is a measure of how much evidence there is against the null hypotheses. The smaller the P-value, the stronger the case is against H0. The general rule is that a small P-value is evidence against the null hypothesis while a large P-value means little or no evidence against the null hypothesis. Please note that little

or no evidence against the null hypothesis is not the same as a lot of evidence for the null hypothesis. Prior to conducting an experiment an α value is set for the experiment. The α value is the value at which the researchers will reject a hypothesis if the P-value is less than α. A typical α value is $\alpha = 0.05$. Sometimes, though, researchers will use a stricter cut-off, $\alpha = 0.01$, or a more liberal cut-off, $\alpha = 0.1$. For all of the experiments in this dissertation, α is set to 0.05. If the computed P-value is equal to or smaller than 0.05, this taken to mean that were the null hypothesis true and an event with probability 1 in 20 or smaller has occurred. Therefore, the two means are statistically different at the $\alpha = 0.05$ level.

All statistical tests consist of the same basic steps:

1. The α value is set before the experiment is conducted.

2. From the experiment data, the value of a test variable is computed.

3. The value of the test variable is compared to a pre-arranged table to find the significance (P-value). The P-value expresses how frequently the test variable would reach the computed value, *if the null hypothesis were true.*

4. The α value is the critical P-value. If computed P is equal to or smaller than α, the P-value is said to be significant.

It is easiest to understand the P-value in a data set that is already at an extreme. Suppose that a drug company alleges that only 50% of all patients who take a certain drug will have an adverse event of some kind. It is believed that the adverse event rate is higher. In a sample of 12 patients, all 12 have an adverse event.

The data supports the belief that the adverse event rate is higher because the data is inconsistent with the assumption of a 50% adverse event rate. Such a result has the same likelihood as flipping a coin 12 times and getting heads each time.

The P-value, the probability of getting a sample result of 12 adverse events in 12 patients assuming that the adverse event rate is 50%, is a measure of this inconsistency. The P-value, 0.000244, is small enough that the hypothesis that the adverse event rate was only 50% would be rejected.

3.1.2   Overview of Studies

The effectiveness of the facial illustration and caricature algorithms on recognition and learning tasks is evaluated in a series of behavioral studies. The hypothesis is: if the facial illustration and caricature algorithms do not affect the recognition speed and accuracy 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 were recognition speed is paramount. To test this hypothesis, three studies were performed that are replications of earlier distinctiveness studies [128]. While these previous studies assessed the effect of human drawn portraits and caricatures on recognition and learning speed, these same studies are used here to validate the computer-generated illustrations and caricaturing techniques. In addition, the computer generated illustrations and caricatures are compared with the source photographs in terms of recognition and learning speeds.

The illustrations used in these studies are computed from source photographs by multiplying thresholded luminance and thresholded brightness images. These illustrations are reminiscent of "line art" images. Therefore, two studies were undertaken to compare the illustrations to images generated using the Caricature generator software and the Canny edge detector. The Caricature generator software is described in Chapter 2.

The Canny edge detector is used in computer vision to create single line weight images from source photographs. The Canny operator works in a multi-stage process. First, the image is smoothed by Gaussian convolution. Then a simple 2-D first derivative operator is applied to the smoothed image to highlight regions of the image with high first spatial derivatives. Luminance edges in the source image give rise to ridges in the gradient magnitude image. The algorithm then tracks along the top of these ridges and sets to zero all pixels that are not actually on the ridge top so as to give a thin line in the output image. This process is known as non-maximal suppression. The tracking process exhibits hysteresis controlled by two thresholds: T1 and T2 with T1 > T2. Tracking can only begin at a point on

Figure 3.1. Examples of photographs and illustrations used as visual stimulus. Left: photographs. Right: facial illustrations. The photographs are part of the Aleix Face Database and are included here courtesy of Aleix Martinez of Purdue University. The entire face database is accessible online [95].

a ridge higher than T1. Tracking then continues in both directions out from that point until the height of the ridge falls below T2. This hysteresis helps to ensure that noisy edges are not broken up into multiple edge fragments [29].

3.2   Recognition Time

This study assesses the recognition time of familiar faces presented as illustrations, caricatures, and photographs. Based on the results obtained with hand drawn images [128] and Caricature Generator images [15, 22, 117], it stands to reason that photographs may be recognized faster than both facial illustrations and caricatures, while caricatures would elicit a faster response time than the facial illustrations [15].

Participants were presented with sequences of images of familiar faces. Examples of stimulus images are shown in Figure 3.1. Each participant was asked to say the name of the person pictured as soon as that persons face is recognized. Reaction times as well as accuracy of the answers were recorded. Images were presented to the participants in three separate conditions: photographs and facial illustrations, photographs and caricatures, or facial illustrations and caricatures. The experiment conducted in this study is the same as the one conducted by Stevenage [128]. The details of this study are presented in Appendix A.1.

The difference between the mean recognition time for photographs and illustrations of familiar faces is not statistically different, hovering around 1.89 seconds.

The caricatures were recognized on average 0.09 seconds slower than photographs. There is no statistical difference, at the $\alpha = 0.05$ level, in the reaction time between caricatures and facial illustrations. In each condition, the accuracy of recognition is higher than 98%, indicating that there is no significant speed for accuracy trade-off in this study.

From this study, it can be concluded 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 aloud during this study when shown caricatures of familiar faces. Laughter toward the stimulus may mean that caricatures can be used for entertainment value in situations where recognition speed is not of the utmost importance.

3.3   Learning Speed

At least one study indicates that caricatures do not improve the ability to learn [58], while others have shown that caricatures of unfamiliar faces are learned more quickly than the same faces shown as fully detailed drawings or as photographs [97, 128]. The hypothesis is that the outcome of these studies is strongly dependent on the particular techniques used to create the stimuli. Because the current approach for creating illustrations and caricatures is very different from those used in previous studies, participants were subjected to a learning task to assess how illustrations and caricatures influence the ability to learn human faces.

In this study, each participant was presented with images of 12 unfamiliar faces in sequence. Each face is verbally assigned a name. Each participant is shown exclusively photographs, facial illustrations, or caricatures. Next, the images were randomly reordered and presented again. The participant is asked to recall the name corresponding to each image. Throughout this process, the study coordinator corrected mistakes and repeated the names that the participant could not remember. The images were shuffled, and the process is repeated until the participant

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

Table 3.1. Learning speed studies, showing the minimum, maximum, and mean number of trial iterations for the study presented in Study 2.

Condition	5 sec.	15 sec.	60 sec.
Train Illustration, Test Illustration	69.32	90.63	98.96
Train Illustration, Test Photograph	57.58	63.10	76.34
Train Photograph, Test Illustration	65.58	73.07	90.63
Train Photograph, Test Photograph	88.64	97.92	98.96

Table 3.2. This table shows the mean accuracy results for the four training and testing methods across three different training times.

could correctly name all 12 faces once without error. The details of this study are presented in Appendix A.2.

Participants learned to recognize 12 unfamiliar faces more than twice as fast in trials with illustrations compared to trials with photographs. On average 2.3 passes were required for illustrations while photographs required an average of 5.4 passes. Participants learned to recognize caricatures of unfamiliar faces 1.5 times faster than with photographs. On average 3.4 passes were required for illustrations while photographs required an average of 5.4 passes. The statistics for the rate of learning (number of trials) for each representation of the faces is shown in Table 3.1.

Three additional studies investigated whether the amount of training time affected recognition accuracy. Participants were presented with photographs and illustrations with different training times, (5, 15, and 60 seconds), before the testing phase of the study. The data collected in these studies suggests that as the amount of training time decreases, photographs may become better than illustrations in face learning tasks. The mean accuracy for the four training conditions and the three training lengths are shown in Table 3.2. From this data, one can conclude

User study	Accuracy		
	Portraits	Caricatures	Photographs
Rhodes et al. [117]	38%	33%	-
Stevenage [128]	96%	100%	-
Current study	99%	98%	98%

Table 3.3. This table shows the recognition accuracy results for facial illustrations, caricatures, and photographs across studies. These results show that illustrations perform better than previous computer drawing methods and may be as accurate as artist drawn images.

that performance in recognition tasks may degrade more rapidly when illustrations are used as stimulus than when photographs are used. Additional studies would need to be made to make a strong statistical argument for this point. The details of these studies are presented in Appendices A.3, A.4 and A.5.

3.4   Learning Accuracy

Participants who were trained using the portrait or caricature images participated in a follow-up study using the original photographs, each shown once. In this study, 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 A.6. The accuracy witnessed with the 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. [117]). The recognition accuracy measured in this study is compared with studies using hand-drawn images, Stevenage [128], and images created using the Caricature Generator [22] in Table 3.3.

The results of this study show that the illustrations produced using the methods reported in this dissertation may be as accurate as artist drawn images. In addition, the results suggest that the illustrations perform far better than images produced using the Caricature generator software. This study also demonstrates that the participants were very accurate when photographs were used as stimulus images.

Figure 3.2. Examples of photographs and Canny edge images used as visual stimulus. Left: facial photographs. Right: Canny edge images. The photographs are part of the Aleix Face Database and are included here courtesy of Aleix Martinez of Purdue University. The entire face database is accessible online [95].

3.5   Photographs versus Canny Edge Detected Images

Another study was conducted to investigate the ability of participants to learn and recognize sets of faces presented as photographs or Canny edge detected images. The Canny operator was designed to be an optimal edge detector [29]. It takes as input a gray scale image, and produces as output an image showing the positions of tracked intensity discontinuities. Examples of stimulus images are shown in Figure 3.2.

The purpose of this study is to enable a comparison between the facial illustration methods presented in this dissertation and the current "state of the art" automatic edge line algorithm. The results of this study show that participants trained with Canny edge images and tested on photographs were accurate 38% of the time. In comparison, participants trained with illustrations and tested on photographs were accurate 63% of the time. Participants trained on photographs and tested on Canny edge images were accurate 59% of the time while Participants trained on photographs and tested on illustrations were accurate 73% of the time. The details of this study are presented in Appendix A.7.

The output of the Canny operator is determined by three parameters: the width of the Gaussian smoothing operator used in the smoothing phase, and the upper and lower thresholds used by the tracker. The Gaussian smoothing operator

Figure 3.3. Examples of photographs and Caricature Generator software images used as visual stimulus. Left: Photographs. Right: Caricature Generator images. The photographs are part of the Aleix Face Database and are included here courtesy of Aleix Martinez of Purdue University. The entire face database is accessible online [95].

is a 2-D convolution operator that is used to blur images and remove detail and noise. Increasing the width of the Gaussian mask reduces the Canny edge detector's sensitivity to noise, at the expense of some fine detail in the image. The localization error in the detected edges also increases slightly as the width of the Gaussian smoothing operator is increased. Usually, the upper tracking threshold can be set quite high and the lower threshold quite low for good results [29]. Setting the lower threshold too high will cause noisy edges to break up. Setting the upper threshold too low increases the number of spurious and undesirable edge fragments appearing in the output. In all of the Canny edge images the width of the Gaussian smoothing operator was set to 0.60, the lower threshold was set to 0.50, and the upper tracking threshold was set to 0.90.

3.6   Photographs versus Caricature Generator Images

This study investigates the ability of participants to learn and recognize sets of faces presented as photographs or Caricature Generator software images. The motivation for this study is the fact that "Caricature Generator" images have become the *de facto* standard for conducting research in face recognition [16, 117–119, 128]. Examples of stimulus images are shown in Figure 3.3.

The results of this study show that the participants were far less accurate when trained to recognize facial images using "Caricature Generator" images and that this result is statistically significant. Participants trained with "Caricature Generator" images and tested on photographs were accurate 30% of the time. In comparison, participants trained with illustrations and tested on photographs were accurate 63% of the time. Participants trained on photographs and tested on "Caricature Generator" images were accurate 39% of the time while Participants trained on photographs and tested on illustrations were accurate 73% of the time. The details of this study are presented in Appendix A.8.

3.7 Functional Magnetic Resonance Imaging

fMRI can be used to identify regions of the brain that are associated with a given stimulus [81, 82, 99, 103, 121, 133]. The most commonly used technique for the localization and measurement of brain activity is based on blood oxygenation level dependent (BOLD) imaging. BOLD fMRI techniques measure changes in the magnetic field within a small volume of tissue resulting from changes in blood oxygenation within the tissue. The MRI signal intensity reflects the concentration of water within the sample and is dependent on the chemical and physical environment in which the water molecules reside [52]. The MRI scanner measures the emission of radio frequency energy by hydrogen atoms in the water molecules of the brain [131]. In the presence of a magnetic field, hydrogen atoms absorb energy applied at a characteristic radio frequency. The hydrogen atoms will then emit energy at the same radio frequency until they gradually return to their original equilibrium state [102]. The MRI scanner measures the sum total of the emitted radio-frequency energy.

BOLD imaging takes advantage of the magnetic properties of the two different types of hemoglobin in the blood: deoxyhemoglobin which is paramagnetic, and oxyhemoglobin which is weakly diamagnetic [108]. Paramagnetic atoms have net orbital or spin magnetic moments that are capable of being aligned in the direction of the applied field. Diamagnetism is the magnetization in the opposite direction

to that of the applied magnetic field. All substances are diamagnetic, but it is a weak form of magnetism and may be masked by other, stronger forces, for instance a magnetic field. Therefore, changes in oxygenation of the blood can be observed as the fMRI signal changes [75, 102, 132]. The signal measured with BOLD fMRI is the ratio of oxyhemoglobin to deoxyhemoglobin in the blood in an area of the brain. Increased blood flow to an area changes the concentration of deoxyhemoglobin in the nearby tissue. The presence of deoxyhemoglobin in the blood vessels causes a darkening of the image in those voxels containing vessels [100, 101] due to a susceptibility difference between the vessel and its surrounding tissue [131].

The emerging model of increased blood flow, called the hemodynamic response, results from an increase in oxygen consumption due to an increase in neuronal activity [28, 65, 88, 99, 102]. fMRI BOLD therefore measures neuronal activity indirectly via an assumed correlation with local blood flow [8]. This assumption, that the fMRI signal is approximately proportional to a measure of local neural activity, is referred to as the linear transform model of the fMRI signal [21]. fMRI BOLD measurements are averaged over several cubic millimeters of tissue and a time period of several seconds allowing for the possibility of a complex nonlinear relationship between a given stimulus and the resulting neuronal activity [70]. While fMRI BOLD imaging via the linear transform model is not a directly quantifiable measure of neuronal activity, it is a useful approximation of the complex interactions between blood flow and neuronal activity [8, 32, 94, 146]. The precise correlation between neuronal activity, metabolic demand, and the hemodynamic response is still not fully defined. However, it has been shown that the fMRI BOLD signal correlates well with visual perception [10].

3.8 Designing the fMRI Study

The design of fMRI studies relies on the ability to detect stimulus evoked signal changes in a series of MRI signals and to extract regions of activation from these signals using statistical techniques [35]. The most common stimulus presentation pattern in fMRI studies is to alternate periods of stimulus and rest. Using this

design, robust activation can be detected in the primary motor, sensory, and visual cortex [90]. These periods are labeled *on* and *off*. This type of experiment is called a boxcar design. The duration of the stimulus and rest periods needs to be long enough to accommodate the hemodynamic response. Therefore, a time period of between eight and sixteen seconds is chosen. The rest and stimulus periods are repeated for as long as necessary to gain enough contrast, compared to noise, to detect the activation. However, the duration of the study is a careful balance between how long the participants can comfortably lay still, the amount of data required to obtain sufficient contrast to noise, and the amount of drift inherent in the scanner. In addition, there is the possibility of the participant habituating to the stimulus causing the BOLD contrast to reduce with time.

The choice of stimulus is critical in the design of fMRI studies. For example, activating the primary visual cortex is straightforward, nearly any visual stimulus will do. However, to determine the regions of the brain responsible for color discrimination is more difficult [141]. It is necessary to design the *on* and *off* periods such that there is only one well defined signal difference between them, which only activates the brain regions responsible for the task. Single region activation is not always possible and so a hierarchy of studies often performed. For example, to identify the brain regions responsible for Task A, a study can be performed that involves Task A and Task B, and then one that only involves Task B. The regions responsible for Task A would presumably be those activated in the first study but not the second. This experimental method assumes that the BOLD signal is a linear system, which may not be the case [99, 133]. In addition, there may be some unaccounted for differences in the *on* and *off* periods that may affect the BOLD signal.

Another problem when dealing with cognitive events such as memory is that a stimulus must be presented and a response given. Many stimuli give better activation if a response is required to be made [60, 126, 142]. Both the stimulus and the response must be compensated for by being included in the *off* period or an additional study must be performed later that involves similar stimulus and

response, but not the cognitive task performed in the original study. Alternatively, the stimulus may be presented in a different modality and the regions common to both stimulus presentation types can be assumed responsible for the cognitive task of interest.

In the case of using illustrations and photographs, the fMRI study needed to be different from the psychophysical studies carried out in the first part of this dissertation because fMRI constrains the experimental methods that can be used. The choice of the optimum parameters for fMRI is always a compromise, and more often depends on what is available than on what is desirable. The following list delineates the changes that had to be made.

1. Boxcar Design of fMRI Studies – fMRI Studies are typically conducted using 8 to 16 second stimulus blocks to contrast blood flow in affected regions of the participants brain [13, 82, 103].

2. Large Number of Stimulation Images – The task that was developed required pilot studies that evaluated photographs, illustrations, and caricatures as stimulus. To avoid repetition a large number of stimulus images were needed.

3. Contrast Image – The block design of the fMRI BOLD studies requires a resting condition to contrast with the active condition.

4. No Verbal Response – In the previous psychophysical studies verbal responses were given by the participants in reaction to the visual stimulus.

5. Active Task – Numerous studies have shown that measurable BOLD response is higher in studies with an active task design than in studies with a passive task [60, 126, 142].

6. No Auditory Stimulus – In the previous psychophysical studies auditory stimulus were given to the participants in at the same time as the visual stimulus.

Figure 3.4. These images with differing facial expressions are part of the Aleix Face Database and are included here courtesy of Aleix Martinez of Purdue University. The entire face database is accessible online [95].

7. Incorporate Behavioral Studies – A perception study was incorporated into the fMRI study to be certain that the state of being inside of the fMRI apparatus did not affect the perceptual phenomena being observed.

3.8.1   Large Number of Stimulation Images

To avoid repetition in the faces viewed by the participants in the fMRI studies a large number of stimulus images were needed. The *Aleix Face Database* [96] was used as a source of facial images. Examples of *Aleix Face Database* images are shown in Figure 3.4. This face database was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the Universitat Autonoma de Barcelona. It contains over 4,000 color images corresponding to 126 faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The pictures were

Figure 3.5. Examples of contrast images used in the behavioral studies and in the fMRI study. Row 1: photograph, illustration, caricature, superportrait illustration. Row 2: Canny edge detected, Caricature Generator software, phase-scrambled illustration, phase-scrambled photograph. The phase-scrambled images were used to create baseline-viewing conditions in pilot versions of the fMRI study. The photograph is part of the Aleix Face Database and is included here courtesy of Aleix Martinez of Purdue University. The entire face database is accessible online [95].

taken at the CVC under strictly controlled conditions. No restrictions on wear (clothes, glasses, make-up, hair style, etc.) were imposed on the participants. Each person participated in two sessions, separated by two weeks (14 days) time. The same poses and conditions were photographed in both sessions. After culling images of individuals with glasses and facial hair from the database, 78 usable images remained.

3.8.2 Contrast Image

Examples of contrast images used in the behavioral studies and in the fMRI study are shown in Figure 3.5. A robust method for generating BOLD contrast is to continually acquire magnetic resonance images while the participants are presented

Figure 3.6. Left: The Lumina LP-400 controller. Right: The Lumina LP-400 response pads. This image is included here courtesy of Hisham Abboud and the Cedrus Corporation [1].

with of sensory stimulation [13, 82, 103]. These blocks of sensory information are inter-spaced with blocks of a resting condition to establish a baseline. The fMRI images acquired during the resting condition are subtracted from the corresponding fMRI images acquired during the stimulation condition to form a difference image. These difference images show BOLD contrast in brain regions where the stimulated neuronal activity leads to a reduction in the concentration of deoxyhemoglobin [100, 103].

3.8.3   Active Task and No Verbal Response

In fMRI verbal response is not viable due to the fact that any head or neck movement on the part of the participants can negatively effect the resulting fMRI scan [14]. Requiring the participant to respond orally almost always results in head movement concurrent with the stimulus. All response movements need to be small to reduce head motion during the scanning.

To conduct a study with an active task but no verbal response the Lumina LP-400 response system was used. The controller and pads are shown in Figure 3.6. The Lumina LP-400 was designed specifically for use in an fMRI environment. The
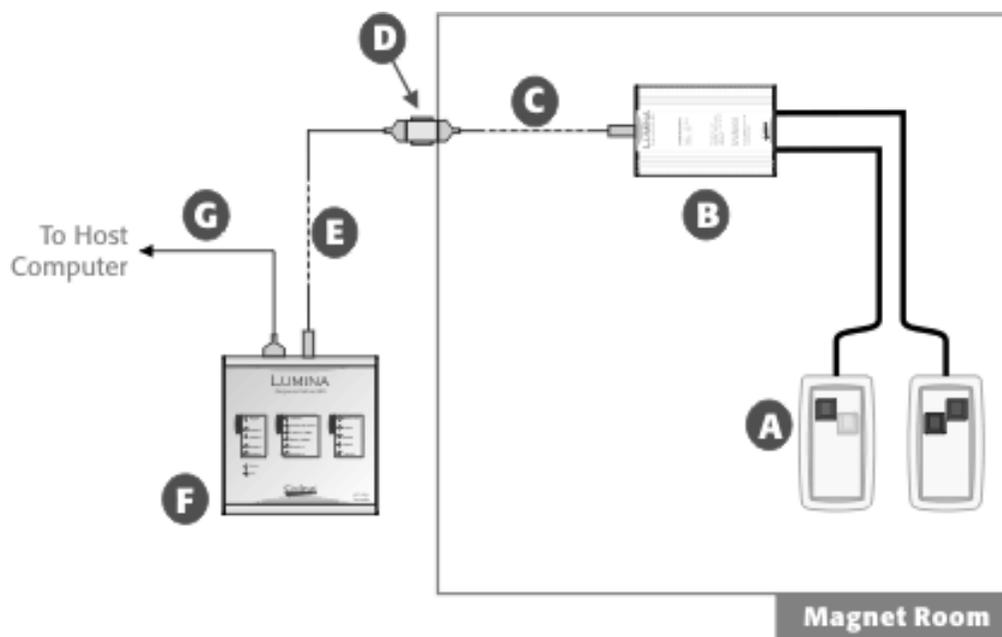
Figure 3.7. A) Response Pads: These are built of 100% plastic and fiber optics. B) OTEC Unit: Converts electricity to light and is connected to the pads via protected fiber optic cables. C) Shielded Cable: Connects the OTEC unit to the penetration panel. Uses wires and connectors of the highest quality. D) RF Filter: Optionally installed if the penetration panel does not have a built-in filter. E) Shielded Cable: Connects the penetration panel to the controller. F) Controller: Detects key presses, times them, performs TTL I/O, and connects to the host computer. G) Serial Cable: Connects the controller to the host computer. This image is included here courtesy of Hisham Abboud and the Cedrus Corporation [1].

system consists of several components including the pads, controller, and over 40 meters of shielded cables. An overview of the entire Lumina LP-400 response system is shown in Figure 3.7.

3.8.4   No Auditory Stimulus

Auditory stimulus is difficult to present in an fMRI study [61]. Rapid imaging scanners are very noisy during operation. Although aural cues can be heard, these cues are not as easily detected by participants as visually presented cues. In addition, detecting activation in the primary auditory cortex is difficult since there is always a large amount of ambient sound during both the *on* and *off* periods. To build an fMRI study with an associated behavioral study that avoids
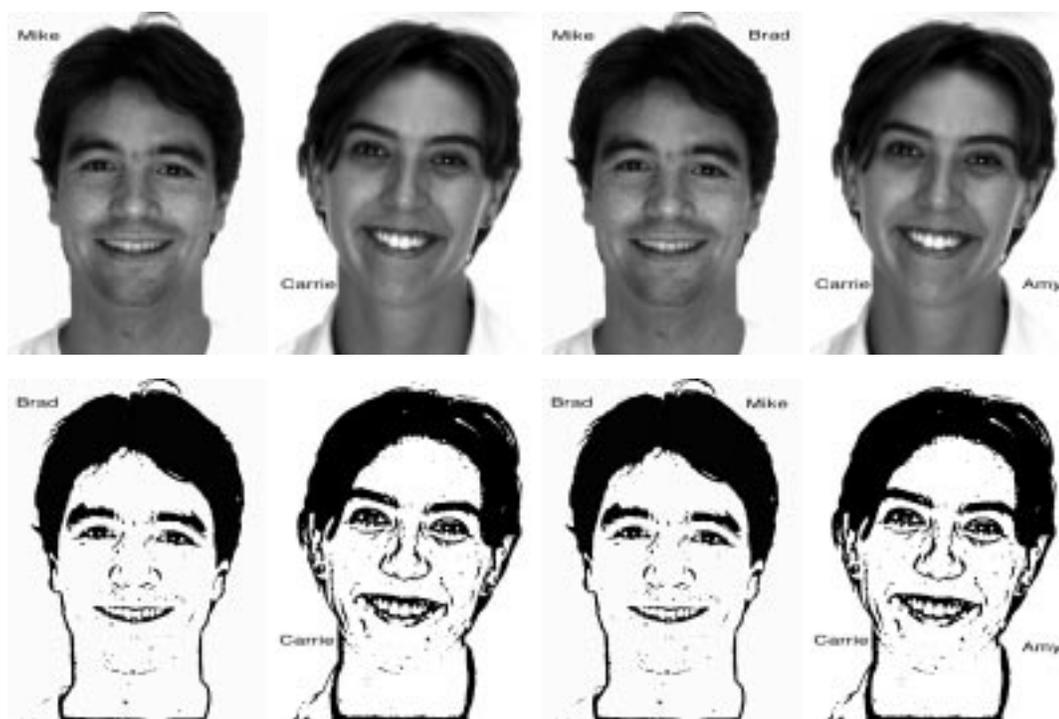
Figure 3.8. Examples of images augmented with name text. These augmented images were used in the fMRI study instead of using an auditory stimulus to identify the faces. Rows 1 and 2: training photographs, testing photographs, training illustrations and testing illustrations. The photographs are part of the Aleix Face Database and are included here courtesy of Aleix Martinez of Purdue University. The entire face database is accessible online [95].

auditory stimulus, the face images were augmented with either one or two first names, depending on whether the images was used for training or testing, printed in Helvetica font. Examples of such augmented images are shown in Figure 3.8.

Two behavioral studies were conducted in which participants were presented with illustrations or photographs augmented with names. The results of these studies show that photographs and illustrations augmented with names are not significantly different as stimulus images in this type of task. In addition, these studies demonstrate that response accuracy is not different from the studies in which an auditory stimulus was given. The full details of these studies are presented in Appendices A.9 and A.10.

Condition	Outside fMRI mean	In fMRI mean	P-value
Train Illustration, Test Illustration	82.51	81.82	0.179
Train Illustration, Test Photograph	72.78	79.80	0.917
Train Photograph, Test Illustration	81.04	81.82	0.336
Train Photograph, Test Photograph	79.55	87.88	0.958

Table 3.4. Recognition study, showing the mean value of correctly identified faces for four methods of training and testing. The first column is in a study outside of the fMRI machine. The second column shows a study using the same stimulus in the same order but was conducted during an fMRI scan.

3.8.5   Incorporate Behavioral Studies

In the fMRI pilot study and the fMRI behavioral study participants were presented with illustrations or photographs augmented with names. The fMRI pilot study was conducted outside of the fMRI scanner under lab conditions. The fMRI study used the same stimulus in the same order as the fMRI pilot. However, the fMRI study took place during an fMRI scan. In the fMRI pilot study participants choose the name that they thought was correct by pressing a designated key on the laptop computer keyboard. In the fMRI behavioral study participants choose the name that they thought was correct by pressing a key on the Lumina response pad while inside the fMRI machine. These two studies show that the results of the fMRI pilot study and the fMRI behavioral study are very similar. None of the means were found to be significantly different. The full details of these studies are presented in Appendices A.11 and A.12. The results of the study showing the mean accuracy of correctly identified faces outside of the fMRI machine and inside the fMRI are shown in Table 3.4.

3.8.6   Illustrations versus Superportraits and Caricatures

Two separate studies were undertaken to investigate the ability of participants to learn and recognize sets of faces presented as illustrations and either superportraits or caricatures. An earlier study, reported in Section 3.2, pointed to the fact that recognition time and accuracy are very similar when using photographs or

Figure 3.9. Examples of illustrations and superportraits used as visual stimulus. Left: facial illustrations. Right: superportraits.

illustrations as stimulus. In addition, the same study showed that illustrations and caricatures elicit similar responses in behavioral tasks.

The results of these studies show that superportraits and caricatures elicit the same response as illustrations in behavioral tasks when using the current stimulus images and experimental procedures. Therefore, caricatures and superportraits would not need to be included in the fMRI behavioral experiment. It was desirable to drop the caricatures and superportraits for two reasons: to simplify the experiment and allow for a greater number of contrasts between the *on* and *off* states of the experiment and to cut the time duration of the fMRI experiment down to 9 minutes to avoid drift in the fMRI machine. Examples of stimulus images are shown in Figures 3.9 and 3.10.

The results of these studies show that the participants were more accurate at recognizing facial images when trained on illustrations than when trained with either superportraits or caricatures. However, the mean accuracy was not significantly different between illustrations and superportraits or between illustrations and caricatures. Therefore, caricature and superportrait images were not used as stimulus in the fMRI study. The details of these studies are presented in Appendices A.13 and A.14.

Figure 3.10. Examples of illustration and caricature images used as visual stimulus. Left: facial illustrations. Right: caricatures.

3.9   fMRI Study Results

The recognition invariance demonstrated in the user studies suggests that different brain structures may be involved in the perception of the illustrations than in the perception of photographs. Statistical parametric maps computed from the fMRI data confirm this hypothesis. Statistical parametric mapping (SPM) is used to identify functionally specialized brain regions and is the most prevalent approach in characterizing functional brain anatomy. SPM is a voxel-based approach that allows the creation of images that show statically significant response to a given stimulus. The details of fMRI data analysis are presented in Appendix B.

Figure 3.11 shows examples of multiple statistical parametric maps computed from the fMRI dataset. The fact that differences occur in the area and level of activation in the decoding images is a new and exciting finding.

The fMRI images at the top of Figure 3.11 were acquired while participants viewed illustrations as stimulus, and show bilateral activation in the frontal lobes of the brain. The fMRI images at the bottom of Figure 3.11, which show activation on only one frontal lobe, were acquired while participants viewed photographs as stimulus. Subjectively it can be observed that the brain images in the lower right corner of Figure 3.11 show a higher overall level of activation as well as activation in different areas that in the other images.
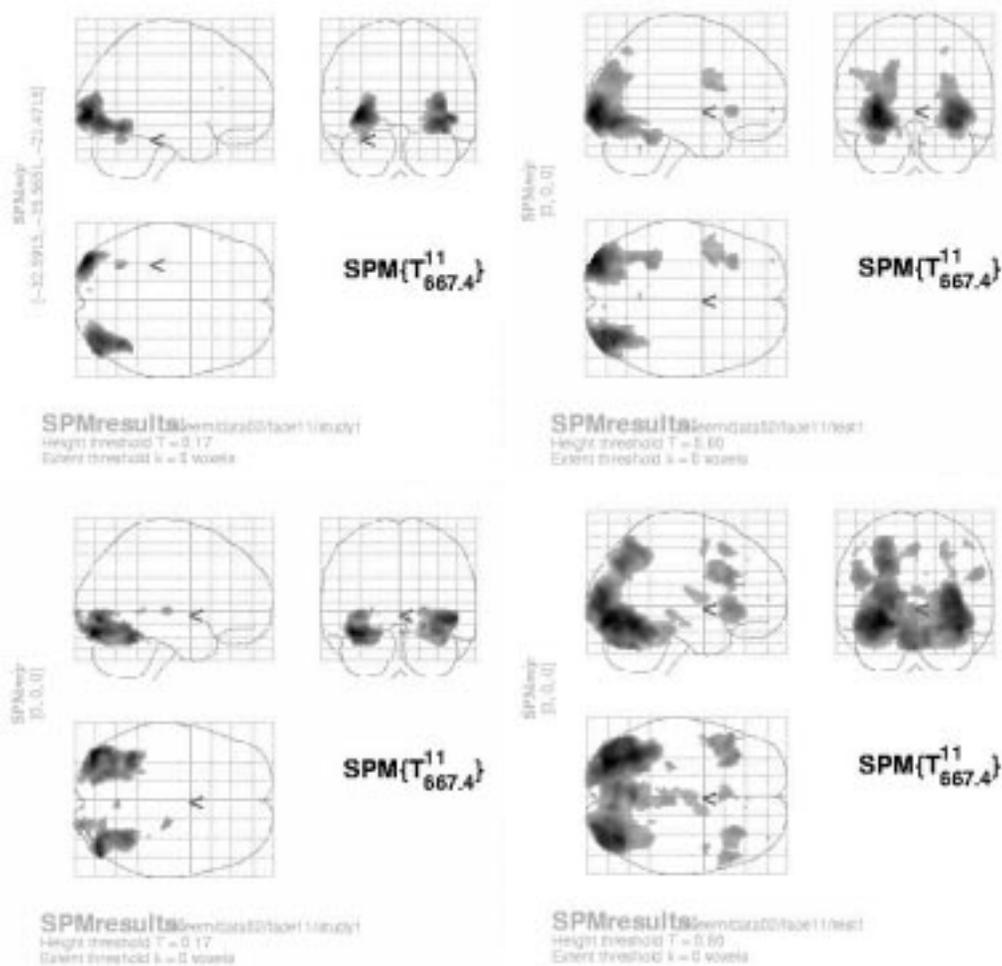
Figure 3.11. These images show a statistical parametric map (SPM) of brain activation during the fMRI experiment. These images are of normalized data over all the participants. Top Row: The SPM on the left shows thresholded activation during the training phase of the fMRI study with photographs as the stimulus (encoding photographs). The SPM on the right shows thresholded activation during the testing phase of the fMRI study with photographs as the stimulus (decoding photographs). Bottom Row: The SPM on the left shows thresholded activation during the training phase of the fMRI study (encoding illustrations) with illustrations as the stimulus. The SPM on the right shows thresholded activation during the testing phase of the fMRI study with illustrations as the stimulus (decoding illustrations).

3.10   Summary

This chapter presented the results of 14 perceptual studies and an fMRI study carried out using illustrations and caricatures,which were created using the methods

of Chapter 2, as stimulus. These studies show that in tasks involving recognition speed the illustrations are not significantly different from photographs as stimulus. In addition, in learning tasks participants presented with illustrations as stimulus performed better than participants presented with photographs. These results are consistent with the results of previous studies were the stimulus images were drawn by professional artists [128]. While these results are interesting, the major impact of these studies is stating to the NPR community that studies in human perception are a viable method of measuring the effectiveness of NPR images. The results of the fMRI study show that the illustrations, which produce the same behavioral results as photographs, produce higher overall activation as well as activation in different areas of the brain.

CHAPTER 4

CONCLUSION

To restate the findings, this dissertation presents four main results:

1. An automated method to produce easily recognizable black-and-white illustrations of human faces from photographs.

2. A user-assisted method to produce caricatures from images of human faces.

3. The results of behavioral studies that show that in some tasks using illustrations yield better user performance than the same tasks using photographs under an objective standard.

4. The results of a functional magnetic resonance imaging (fMRI) study indicating that illustrations and photographs may be processed differently when viewed by humans.

The algorithms presented in Chapter 2 produce easily recognizable illustrations of human faces from photographs while minimizing the amount of computer memory required to encode these images. In addition, Chapter 2 reported on an interactive technique for creating caricatures from facial images. This technique allows non-skilled users to quickly and easily create caricatures.

The behavioral studies documented in Chapter 3 demonstrate that substituting illustrations for fully detailed photographs has no significant impact on recognition speed. In addition, learning tasks can be accomplished faster when illustrations or caricatures are used as stimulus instead of photographs. The results shown in Chapter 3 are consistent with the results of previous studies were the stimulus images were drawn by professional artists [128]. In addition, the illustrations

performed better in accuracy studies than previous computer generated line art methods [15, 22, 117] and Canny edge detected images [29]. While these results are interesting, the major impact of these studies is stating to the NPR community that studies in human perception are a viable method of measuring the effectiveness of NPR images.

The learning speedup and the recognition invariance demonstrated in the behavioral studies Chapter 3 suggest that different brain structures may be involved in the perception of artistic images than are involved in the perception of photographs. In the case of the facial illustrations generated using the methods of this dissertation this seems to be true. The results of the fMRI study show that the illustrations, which produce the same behavioral results as photographs, produce higher overall activation as well as activation in different areas of the brain.

4.1   Future Work

There are many possible future directions for the facial illustration work. Animating human facial illustrations is a natural extension of this work. Processing a series of images using the techniques would lead to an animated portrait. Currently only an image filtering technique is used to produce facial illustrations. Another possibility is to incorporate stroke based techniques to produce images more reminiscent of drawings. In addition, the conditions under which photographs are taken may influence the output of the illustration 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 participants 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 the algorithm against these factors.

The illustrations presented in this work open the door for a number of potential applications. First, the resulting illustrations are suitable for rapid transmission over low bandwidth networks. The illustrations may be stored with only one bit per pixel and image compression may further reduce memory requirements. As

such, one can foresee applications in telecommunications where users may transmit illustrated signature images of themselves in lieu of a caller identification number when making phone calls. While wireless hand-held digital devices are already quite advanced, bandwidth for rapidly transmitting images is still a bottleneck. The caricatures produced using the methods of this dissertation have largely similar characteristics in terms of recognition and learning tasks and can therefore be applied in entertainment oriented applications without significantly impeding task performance. Behavioral studies involving other types of NPR image types seems certain to uncover additional tasks that can be performed better using NPR images as stimulus instead of photographs.

Research in face recognition may also benefit from using the techniques reported in this work. The method for producing superportraits described in Chapter 2 constitutes a viable alternative for further research into this phenomenon, and may prove to be a good replacement for the "Caricature Generator" algorithm that is regarded as the *de facto* standard.

The 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 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? These are intriguing questions that further work in this area might pursue.

Currently a renaissance is occurring in the area of brain function mapping using fMRI. NPR images may have a larger roll to play in the unfolding process of human understanding of the inner workings of the brain. NPR images are simplified versions of reality. By comparing how the brain reacts to NPR images compared to how the brain reacts to photographs scientists may be able gain insight into human visual perception.

APPENDIX

PERCEPTION STUDIES

As stated in Chapter 3, for all of the studies in this dissertation, α is set to 0.05. P-values are computed using two-way analysis of variance (ANOVA). Therefore, if the computed P-value is equal to or smaller than 0.05, this taken to mean that were the null hypothesis true and an event with probability 1 in 20 or smaller has occurred. Therefore, the two means are taken to be statistically different at the $\alpha = 0.05$ level.

A.1   Study 1: Recognition Speed

A.1.1   Participants

In this study, 42 graduate students, postgraduates and research staff acted as volunteers.

A.1.2   Materials

In this study, 60 images depicting the faces of 20 colleagues of the volunteers were used as visual stimuli. Each face was depicted as a gray-scale photograph, an illustration, and a caricature. The photographs were taken using a Kodak 330 digital camera with the flash on. 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 background of the monitor 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.

A.1.3   Procedure

Three two-part studies were conducted, each with 14 Participants. The first part of the study allowed participants to rate their familiarity with a list of 20 names on a seven-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 persons 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 [128, 147, 148].

In the second part of this study, the 12 highest rated faces are selected for each participant and were shown in two of three possible conditions. Participants in Study 1.a saw photographs and illustrations. Study 1.b consisted of photographs and caricatures, and Study 1.c consisted of illustrations and caricatures. The written instructions for this part were: "In this study you will be shown pictures of peoples faces you may know. Each picture will be shown for five seconds followed by a five second interval. Please say the name of each person as soon as you recognize this person." The study coordinator 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 study coordinator recorded the accuracy of the answers and a second study coordinator, who pressed a key at the end of the response, recorded the response time for each image. This stopped the timer that was started automatically upon display of the image.

A.1.4   Results

Participants 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

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

Table A.1. Recognition speed results, showing the minimum, maximum, and mean time for each condition in each study.

illustrations compared to caricatures ($p = 0.07$). The accuracy for recognizing photos, illustrations and caricatures are 98%, 99% and 98% respectively. Table A.1 provides minimum, maximum, and mean times recorded for each condition on each study.

A.2   Study 2: Learning Speed

A.2.1   Participants

In this study, 30 University of Utah graduate students, postgraduates and research staff acted as volunteers. They were selected for unfamiliarity with the faces presented in this study.

A.2.2   Materials

Gray-scale photographs of the faces of six males and six females were used as stimulus in this study. An identical pilot study as in Study 1 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"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 1970s (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.

A.2.3   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 exclusively with photographs, illustrations or caricatures. Each participant was first shown all 12 faces, one image at a time, for about three seconds and was told the name assigned to that face. The faces were then shuffled and individually presented to the participant 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 study.

A.2.4   Results

Illustrations were learned substantially more quickly than photos ($p < 0.001$). In this trial, neither caricatures versus photos nor illustrations versus caricatures could not be distinguished statistically ($p = 0.093$, $p = 0.081$). Learning appears to be quickest when the faces were presented as illustrations, followed by caricatures and then photographs.

A.3   Study 3: Illustrations versus Photographs with a Sixty Second Learning Time

This study investigates the ability of participants to learn and recognize sets of faces presented as illustrations or photographs.

A.3.1 Participants

In this study, 12 University of Utah undergraduates participated as part of a research extra credit opportunity for their psychology class. All participants were tested individually, none knew of the hypothesis being tested.

A.3.2 Materials

In this study, 64 face images from the AR Face Database [96], 32 males and 32 females, were used as stimulus. The images were presented to the participants as gray-scale photographs or as illustrations. An identical pilot study as in Study 1 was carried out and the 32 facial illustrations, 16 males and 16 females, were all rated as being a good likeness. Facial images were presented to the participants using a Macintosh Ibook laptop computer at a distance of 24 inches. The background of the laptop monitor was set to black and displayed images that subtended a visual angle of 12.9 degrees. The participants rated each face using a 5-point Likert-type scale on paper. Likert scales are commonly used in attitudinal measurements. This type of scale uses a five-point scale ranging from strongly agree, agree, neither agree nor disagree, disagree, strongly disagree to rate people's attitudes. Variants of the Likert-scale exist that use any number of points between three and ten. Four different rating scales were used by the participants to rate the face based on the faces: attractiveness, distinctiveness, honesty and likability.

A.3.3 Procedure

This study was broken into two phases, a learning phase and a test phase. In the learning phase, each face was presented to the participant for 15 seconds and rated by the participant on each of four characteristics: attractiveness, distinctiveness, honesty and likability, for a total study time of one minute. Thirty-two faces were shown to each participant in the learning phase of the study. Of the 32 faces presented to each participant in the learning phase of the study, 16 were presented

as photographs and 16 were presented as illustrations. When all of the ratings were complete, participants began the test phase of the study, in which they were shown that same set of 32 faces that they had rated in the learning phase along with a set of 32 new faces that they had not seen before, for a total of 64 faces. The faces were intermixed in the testing phase. The participants' task was to determine which faces they had seen before (an old face) and which ones they had not seen before (a new face). Of the 32 new faces, 16 were presented as photographs and 16 were presented as line drawings. Out of the 32 old faces, the 16 that were presented as photographs, 8 of them changed modality (e.g. 8 of them were presented as illustrations in the test phase) and 8 remained the same (e.g. 8 remained photographs). Throughout this study, men and women always represented an equal number of faces in all conditions.

A.3.4   Design

1. Participants were given a consent form to read and sign.

2. Participants were then told that they would be participating in a study that would examine memory for faces.

3. Participants were then given a detailed description of the learning phase of the study. Participants were then read the following verbal instructions: "You are going to be presented with a series of faces, you will be asked to study each face and will be tested on your memory for these faces later. The faces you see are comprised of line drawings and photographs. For each face you see, you will be asked to rate that face on its attractiveness, distinctiveness, honesty and likability. Use the time the face is presented to study the face carefully. After all of the ratings are complete, you will begin the test phase of the study which I will explain then."

4. Participants were then given a sheet of paper with a rating scale from to five for each of the four characteristics: attractiveness, distinctiveness, honesty and likability.

5. After the participant had rated each of the 32 facial images for each of the four characteristics, the participant was read the following instructions. "In the second phase of the study, you will be shown those same faces along with a set of new faces. You will be asked to determine which faces you see are old faces (faces you were asked to rate) and which faces are new faces (faces you were not asked to rate). Some of the faces that you were shown in the first phase will be presented in the second phase in a different form. For example, you may be shown this face in the ratings phase of the study (show photo) and be shown this face in the test phase of the study (press any key to show line drawing of same face). You would say that this face is old since it is the same face, even though it is in a different format. It may also happen in the opposite way, where you may be shown the line drawing in the first phase, and be shown the photograph in the second phase, still this would be considered and old face. Press the z key if you have seen the face before (old) and the m key if the face is a new face. Make your decision as quickly as possible. Do you have any questions?"

6. The participants were then presented with the 64 faces and made judgments about which faces were new and which were new. The participants registered their judgments via marked keys on the laptop computer keyboard.

In the learning phase of the study, 32 face images were shown to each participant. The images were mixed and the breakdown is as follows:

16 gray-scale photographs

   8 male

   8 female

16 illustrations

    8 male

    8 female

    In the testing phase of the study, 64 face images were shown to each participant. The images were remixed from the learning phase and the breakdown is as follows:

32 previously seen (old faces)

    16 gray-scale photographs

        8 that changed modality (photograph to illustration)

            4 male

            4 female

        8 that remained in the same modality (photograph)

            4 male

            4 female

    16 illustrations

        8 that changed modality (illustration to photograph)

            4 male

            4 female

        8 that remained in the same modality (illustration)

            4 male

            4 female

32 not previously seen (new faces)

    16 gray-scale photographs

        8 male

        8 female

Condition	Min	Max	Mean	Std. Dev.
Train Illustration, Test Illustration	88	100	98.96	3.61
Train Illustration, Test Photograph	57	100	76.34	12.49
Train Photograph, Test Illustration	38	100	90.63	18.56
Train Photograph, Test Photograph	88	100	98.96	3.61

Table A.2. This table shows the minimum, maximum, and mean percentage of correctly identified faces for the four methods of training and testing. Participants were trained for sixty seconds with each image.

16 illustrations

8 male

8 female

A.3.5   Results

Descriptive statistics for the rate of learning over the four learning and testing conditions are shown in Table A.2. In this study, the mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on photographs and tested on photographs were not statistically different ($p = 1.0$). The mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on illustrations and tested on photographs were statistically different at the $\alpha = 0.05$ level ($p < 0.001$). The mean percentage accuracy of participants trained on photographs and tested on illustrations versus those of participants trained on photographs and tested on photographs were not statistically different ($p = 0.166$). New photographs were recognized as new faces 90 percent of the time and new illustrations were recognized as new faces 93 percent of the time. The mean accuracy for recognizing new faces from photographs was not significantly different from the mean accuracy of recognizing new faces from illustrations ($p = 0.464$).

A.4 Study 4: Illustrations versus Photographs with a Fifteen Second Learning Time

A.4.1 Participants

In this study, 12 University of Utah undergraduates participated in the study. All participants were tested individually, none knew of the hypothesis being tested.

A.4.2 Materials

The materials used in this study were the same as those reported in Appendix A.3.

A.4.3 Procedure

The procedure followed in this study was the same as that reported in Appendix A.3, with the exception that the time period that each stimulus image was shown was shortened.

A.4.4 Design

The design of this study was the same as that reported in Appendix A.3.

A.4.5 Results

Descriptive statistics for the rate of learning over the four learning and testing conditions are shown in Table A.3. In this study, the mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on photographs and tested on photographs were not statistically different ($p = 0.131$). The mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on illustrations and tested on photographs were statistically different at the

Condition	Min	Max	Mean	Std. Dev.
Train Illustration, Test Illustration	63	100	90.63	12.07
Train Illustration, Test Photograph	13	100	63.10	27.99
Train Photograph, Test Illustration	50	100	73.07	16.34
Train Photograph, Test Photograph	75	100	97.92	7.21

Table A.3. This table shows the minimum, maximum, and mean percentage of correctly identified faces for the four methods of training and testing. Participants were trained for fifteen seconds with each image.

$\alpha = 0.05$ level ($p = 0.010$). The mean percentage accuracy of participants trained on photographs and tested on illustrations versus those of participants trained on photographs and tested on photographs were statistically different at the $\alpha = 0.05$ level ($p = 0.002$). New photographs were recognized as new faces 81 percent of the time and new illustrations were recognized as new faces 69 percent of the time. The mean accuracy for recognizing new faces from photographs was not different from the mean accuracy of recognizing new faces from illustrations ($p = 0.583$).

A.5 Study 5: Illustrations versus Photographs with a Five Second Learning Time

A.5.1 Participants

In this study, 12 University of Utah undergraduates participated in the study. All participants were tested individually, none knew of the hypothesis being tested.

A.5.2 Materials

The materials used in this study were the same as those reported in Appendix A.3.

A.5.3 Procedure

The procedure followed in this study was the same as that reported in Appendix A.4.

A.5.4 Design

The design of this study was the same as that reported in Appendix A.3.

A.5.5 Results

Descriptive statistics for the rate of learning over the four learning and testing conditions are shown in Table A.4. In this study, the mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on photographs and tested on photographs were statistically different at the $\alpha = 0.05$ level ($p = 0.029$). The mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on illustrations and tested on photographs were not statistically different ($p = 0.231$). The mean percentage accuracy of participants trained on photographs and tested on illustrations versus those of participants trained on photographs and tested on photographs were statistically different at the $\alpha = 0.05$ level ($p = 0.001$). New photographs were recognized as new faces 72 percent of the time and new illustrations were recognized as new faces 60 percent of the time. The mean accuracy for recognizing new faces from photographs was not significantly different from the mean accuracy of recognizing new faces from illustrations ($p = 0.688$).

A.6 Study 6: Learning Accuracy

A.6.1 Participants

In this study, 20 of the participants from Study 2, those who were presented with illustrations or caricatures, took part.

Condition	Min	Max	Mean	Std. Dev.
Train Illustration, Test Illustration	13	100	69.32	25.84
Train Illustration, Test Photograph	25	88	57.58	20.57
Train Photograph, Test Illustration	13	100	65.58	23.66
Train Photograph, Test Photograph	63	100	88.64	13.06

Table A.4. This table shows the minimum, maximum, and mean percentage of correctly identified faces for the four methods of training and testing. Participants were trained for five seconds with each image.

A.6.2 Materials

Gray-scale photographs of the faces of six males and six females were used as stimulus images in this study. An identical pilot study as in Study 1 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"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 1970s (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.

A.6.3 Procedure

This study explored whether caricatures and illustrations result in a difference in learning accuracy. After participating in Study 2, participants were shown 12 photographs in random order and were asked to recall the associated names.

A.6.4 Results

The number of incorrectly identified faces was recorded for each participant. In both cases, either portrait or caricature training, the participants accuracy was 98%. Hence, at the $\alpha = 0.05$ level, there was no measurable difference in accuracy between participants trained with illustrations or caricatures.

A.7 Study 7: Photographs versus Canny Edge Images

A.7.1 Participants

In this study, 12 University of Utah undergraduates participated. All participants were tested individually, none knew of the hypothesis being tested.

A.7.2 Materials

In this study, 64 face images from the AR Face Database [96], 32 males and 32 females, were used as stimulus. Using these 64 images from the AR Face Database, 64 Canny edge facial images were created using an implementation of the Canny edge detector. The images were presented to the participants as gray-scale photographs or as Canny edge images. Other than the difference in stimulus images the materials used in this study were the same as those reported in Appendix A.3.

A.7.3 Procedure

The procedure followed in this study was the same as that reported in Appendix A.3.

A.7.4 Design

The design of this study was the same as that reported in Appendix A.3, with Canny edge images used instead of illustrations.

A.7.5 Results

Descriptive statistics for the rate of learning over the four learning and testing conditions are shown in Table A.5. In this study, the mean percentage accuracy of participants trained on photographs and tested on photographs versus those of participants trained on CE images and tested on CE images were statistically

Condition	Min	Max	Mean	Std. Dev.
Train Photograph, Test Photograph	62.5	100	85.42	11.72
Train Photograph, Test Canny edge	25	100	59.38	20.03
Train Canny edge, Test Photograph	0.0	71.2	38.82	25.17
Train Canny edge, Test Canny edge	50	100	77.1	16.7

Table A.5. This table shows the minimum, maximum, and mean percentage of correctly identified faces for the four methods of training and testing. Participants were trained with each image for fifteen seconds.

different at the $\alpha = 0.05$ level ($p = 0.002$). The mean percentage accuracy of participants trained on photographs and tested on photographs versus those of participants trained on photographs and tested on CE images were not statistically different ($p = 0.751$). The mean percentage accuracy of participants trained on CE images and tested on photographs versus those of participants trained on CE images and tested on CE images were statistically different at the $\alpha = 0.05$ level ($p = 0.002$).

A.8 Study 8: Photographs versus Caricature Generator Images

A.8.1 Participants

In this study, 12 University of Utah undergraduates participated. All participants were tested individually, none knew of the hypothesis being tested.

A.8.2 Materials

In this study, 64 face images from the AR Face Database [96], 32 males and 32 females, were used as stimulus. Using these 64 images from the AR Face Database, 64 Caricature Generator images, 32 males and 32 females, were constructed with a Java application of the Caricature Generator software. The images were presented to the participants as gray-scale photographs or as Canny edge images. An identical pilot study as in Study 1 was carried out and the 32 Caricature Generator images,

Condition	Min	Max	Mean	Std. Dev.
Train Photograph, Test Photograph	62.5	100	85.1	12.07
Train Photograph, Test Caricature Generator	30	80	38.5	18.8
Train Caricature Generator, Test Photograph	0.0	62.5	30.21	22.90
Train Caricature Generator, Test Caricature Generator	25	87.5	67.0	23.40

Table A.6. This table shows the minimum, maximum, and mean percentage of correctly identified faces for the four methods of training and testing. Participants were trained with each image for fifteen seconds.

16 males and 16 females, were all rated as being a good likeness. Other than the difference in stimulus images the materials used in this study were the same as those reported in Appendix A.3.

A.8.3 Procedure

The procedure followed in this study was the same as that reported in Appendix A.4.

A.8.4 Design

The design of this study was the same as that reported in Appendix A.3, with Caricature Generator images used instead of illustrations.

A.8.5 Results

Descriptive statistics for the rate of learning over the four learning and testing conditions are shown in Table A.6. In this study, the mean percentage accuracy of participants trained on photographs and tested on photographs versus those of participants trained on Caricature Generator images and tested on Caricature Generator images were statistically different at the $\alpha = 0.05$ level ($p = 0.037$). The mean percentage accuracy of participants trained on photographs and tested on photographs versus those of participants trained on photographs and tested on

Caricature Generator images were statistically different ($p < 0.001$). The mean percentage accuracy of participants trained on Caricature Generator images and tested on photographs versus those of participants trained on Caricature Generator images and tested on Caricature Generator images were statistically different ($p = 0.001$).

A.9 Study 11: Name-face Training with Feedback using Twelve Faces

A.9.1 Participants

In this study, 12 University of Utah undergraduates participated in the study. All participants were tested individually, none knew of the hypothesis being tested.

A.9.2 Materials

In this study, 12 face images from the AR Face Database [96], 6 males and 6 females, were used as stimulus. The images were presented to the participants as gray-scale photographs or as illustrations. An identical pilot study as in Study 1 was carried out and the facial illustrations were all rated as being a good likeness. Facial images were presented to the participants using a Macintosh Ibook laptop computer at a distance of 24 inches. The background of the laptop monitor was set to black and displayed images that subtended a visual angle of 12.9 degrees.

A.9.3 Procedure

This study was broken into four phases, a learning phase, a training phase and two testing phases. In the learning phase, each face was presented to the participant for five seconds. Twelve faces were shown to each participant in the learning phase of the study. The faces were intermixed in between each phase. During the training phase of the study participants were presented with each face with two names per face for five seconds. The participants chose the name that they thought was correct

by pressing a key on the laptop computer keyboard. Based on the validity of their response the participant was then shown feedback image that read either correct or incorrect for five seconds. During the training phase, each image was shown to each participant twice. During the first testing session, participants were presented with each face with two names in the same modality, illustrations or photographs, with which the participant had been trained for five seconds. The participants chose the name that they thought was correct by pressing a key on the laptop computer keyboard. In the second training phase of this study, the participants received no feedback about the correctness of their answers. Six participants were trained with photos, and six with lines.

A.9.4 Design

1. Participants were given a consent form to read and sign.

2. Participants were then told that they would be participating in a study that would examine memory for faces.

3. The participants were read the following instructions. "In this study, you will be introduced to a set of people by seeing their face and their name presented on the screen. After you study the faces once, you will be shown a face with two names and you must try to choose the correct name. You will receive feedback about whether you are correct or incorrect. After the training with feedback, you will be presented with each face and two names and you should try to choose the correct name. You will not receive feedback during this test. Finally, you will see the same set of faces in a different modality (e.g. illustrations if you originally studied photos). Try to choose the correct name and again, you will not receive feedback. Choose the correct name by pressing either the z key for the name on the left or the m key for the name on the right. Each face is presented for a limited time, so try to answer while the

face is on the screen. If you do not answer in time, just proceed to the next face. Do you have any questions?"

4. Learning Session: Participants were then allowed to study 12 faces, shown to them as either illustrations or photographs, for five seconds each.

5. Training Session: Participants were presented with each face with two names for five seconds. Participants chose the name they thought was correct and received feedback about the correctness of their answer. During the first training session, each face was presented twice.

6. First Testing Session: Participants were presented with each face with two names for five seconds. Participants chose the name they thought was correct. During the first testing session, each face was presented twice.

7. Second Testing Session: Participants were presented with each face with two names for five seconds. In this testing session, each face was presented in the opposite modality as in the previous sessions. Participants chose the name they thought was correct. During the second testing session, each face was presented twice.

A.9.5 Results

Descriptive statistics for the rate of learning over the four learning and testing conditions are shown in Table A.7. In this study, the mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on photographs and tested on photographs were not statistically different ($p = 0.690$). The mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on illustrations and tested on photographs were not statistically different ($p = 0.919$). The mean percentage accuracy of participants trained on photographs and tested on illustrations versus those of participants trained on photographs and tested on photographs were not statistically different ($p = 0.431$). The mean percentage

Condition	Min	Max	Mean	Std. Dev.
Train Illst, Test Illst	55	92	80.87	12.99
Train Illst, Test Photo	50	100	80.41	13.40
Train Photo, Test Illst	50	92	77.78	13.37
Train Photo, Test Photo	50	100	83.14	13.73
Train Illst, 1st Test Illst, 2nd Test Photo	45	100	77.02	16.09
Train Illst, 1st Test Photo, 2nd Test Photo	42	100	80.93	16.41
Train Photo, 1st Test Illst, 2nd Test Illst	58	92	76.89	9.55
Train Photo, 1st Test Photo, 2nd Test Illst	56	100	80.64	12.59

Table A.7. This table shows the minimum, maximum, and mean percentage of correctly identified faces for eight methods of training and testing. In this table the abbreviation Illst is used for the word illustration.

accuracy of participants trained on illustrations and tested on illustrations then tested a second time on photographs versus those of participants trained on photographs and tested on photographs then tested a second time on illustrations were not statistically different ($p = 0.607$). The mean percentage accuracy of participants trained on illustrations and tested on illustrations then tested a second time on photographs versus those of participants trained on illustrations and tested on photographs then tested a second time on photographs were not statistically different ($p = 0.454$). The mean percentage accuracy of participants trained on photographs and tested on illustrations tested a second time on illustrations versus those of participants trained on photographs and tested on photographs tested a second time on illustrations were not statistically different ($p = 0.450$).

A.10   Study 12: Name-face Training with Feedback using Thirty-Two Faces

A.10.1   Participants

In this study, 12 University of Utah undergraduates participated in the study. All participants were tested individually, none knew of the hypothesis being tested.

A.10.2 Materials

In this study, 32 face images from the AR Face Database [96], 16 males and 16 females, were used as stimulus. The images were presented to the participants as gray-scale photographs or as illustrations. An identical pilot study as in Study 1 was carried out and the facial illustrations were all rated as being a good likeness. Facial images were presented to the participants using a Macintosh Ibook laptop computer at a distance of 24 inches. The background of the laptop monitor was set to black and displayed images that subtended a visual angle of 12.9 degrees.

A.10.3 Procedure

This study was broken into two phases, a training phase and a testing phase. In the learning phase, each face was presented to the participant for five seconds and each face was presented three times. In the learning phase of the study, 32 faces were shown to each participant. During the training phase of the study participants were presented with each face with two names for five seconds and each face was shown three times. The participants choose the name that they thought was correct by pressing a key on the laptop computer keyboard. During the testing phase, the participants were presented with each face with two names in the same modality, illustrations or photographs, with which the participant had been trained, for five seconds. The participants chose the name that they thought was correct by pressing a key on the laptop computer keyboard or pressed another key if they thought that this was a face that they had not previously seen. The modality, illustration or photograph, and the order of presentation were mixed between the training and testing phases of the study.

A.10.4 Design

1. Participants were given a consent form to read and sign.

2. Participants were then told that they would be participating in a study that would examine memory for faces.

3. The participants were read the following instructions. "I am going to introduce you to a group of people by presenting their names with their faces. Please try to remember the name of each person. You will have three study periods where you see each of the faces. After the study period, you will be presented with the faces with two different names, in same format as you saw them (photos) a different format (illustrations), or you will see new faces that you have not seen before. If the face is a new face, press the m key for new. If the face is one that you have seen before (either as a photo or a line drawing), choose the correct name by pressing either the z button for the left name, or the x button for the right name. First, we will show you all of the faces three times. You may rest between study sessions and press any key when you are ready to begin again. Do you have any questions?"

4. Learning Session: Participants were then allowed to study thirty-two faces, shown to them as either illustrations or photographs, for five seconds each. Each face was shown three times.

5. Testing Session: Participants were presented with each face with two names for ten seconds. Participants chose the name they thought was correct using a key press on the laptop computer, or pressed another key to demonstrate that they believed that they had not seen the face before.

A.10.5 Results

Descriptive statistics for the rate of learning over the four learning and testing conditions are shown in Table A.8. In this study, the mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on photographs and tested on photographs were not statistically different ($p = 0.256$). The mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on

Condition	Min	Max	Mean	Std. Dev.
Train Illustration, Test Illustration	38	88	67.11	18.78
Train Illustration, Test Photograph	25	100	57.89	18.67
Train Photograph, Test Illustration	50	86	62.80	13.10
Train Photograph, Test Photograph	43	100	73.36	15.72

Table A.8. This table shows the minimum, maximum, and mean percentage of correctly identified faces for the four methods of training and testing.

illustrations and tested on photographs were not statistically different ($p = 0.168$). The mean percentage accuracy of participants trained on photographs and tested on illustrations versus those of participants trained on photographs and tested on photographs were not statistically different at the $\alpha = 0.05$ level ($p = 0.074$). New photographs were recognized as new faces 72 percent of the time and new illustrations were recognized as new faces 61 percent of the time. The mean accuracy for recognizing new faces from photographs was significantly different from the mean accuracy of recognizing new faces from illustrations at the $\alpha = 0.05$ level ($p = 0.106$).

A.11 Study 13: fMRI Pilot, Name-face Training without Feedback using Twelve Faces and Intermixed Blocks of Illustrations and Photographs

A.11.1 Participants

In this study, 12 University of Utah undergraduates participated. All participants were tested individually: none knew of the hypothesis being tested.

A.11.2 Materials

In this study, 12 face images from the AR Face Database [96], 6 males and 6 females, were used as stimulus. The images were presented to the participants as gray-scale photographs or as illustrations. An identical pilot study as in Study 1 was carried out and the facial illustrations were all rated as being a good likeness.

Facial images were presented to the participants using a Macintosh Ibook laptop computer at a distance of 24 inches. The background of the laptop monitor was set to black and displayed images that subtended a visual angle of 12.9 degrees.

A.11.3   Procedure

This study was broken into two phases, a training phase and a testing phase. In the learning phase, each face was presented to the participant for five seconds and each face was presented three times. In the learning phase of the study, 32 faces were shown to each participant. During the training phase of the study participants were presented with each face with two names for five seconds and each face was shown three times. The participants choose the name that they thought was correct by pressing a key on the laptop computer keyboard. During the testing phase, the participants were presented with each face with two names in the same modality, illustrations or photographs, with which the participant had been trained, for five seconds. The participants chose the name that they thought was correct by pressing a key on the laptop computer keyboard or pressed another key if they thought that this was a face that they had not previously seen. The modality, illustration or photograph, and the order of presentation were mixed between the training and testing phases of the study.

A.11.4   Design

1. Participants were given a consent form to read and sign.

2. Participants were then told that they would be participating in a study that would examine memory for faces.

3. The participants were read the following instructions. "I am going to introduce you to a group of people by presenting their names with their faces. Please try to remember the name of each person. You will have three study

Condition	Min	Max	Mean	Std. Dev.
Train Illustration, Test Illustration	67	100	82.51	12.03
Train Illustration, Test Photograph	42	100	72.78	17.71
Train Photograph, Test Illustration	55	92	81.04	11.36
Train Photograph, Test Photograph	55	100	79.55	79.55

Table A.9. This table shows the minimum, maximum, and mean percentage of correctly identified faces for the four methods of training and testing.

periods where you see each of the faces. After the study period, you will be presented with the faces with two different names, in same format as you saw them (photos) a different format (illustrations), or you will see new faces that you have not seen before. If the face is a new face, press the m key for new. If the face is one that you have seen before (either as a photo or a line drawing), choose the correct name by pressing either the z button for the left name, or the x button for the right name. First, we will show you all of the faces three times. You may rest between study sessions and press any key when you are ready to begin again. Do you have any questions?"

4. Learning Session: Participants were then allowed to study thirty-two faces, shown to them as either illustrations or photographs, for five seconds each. Each face was shown three times.

5. Testing Session: Participants were presented with each face with two names for ten seconds. Participants chose the name they thought was correct using a key press on the laptop computer, or pressed another key to demonstrate that they believed that they had not seen the face before.

A.11.5 Results

Descriptive statistics for the rate of learning over the four learning and testing conditions are shown in Table A.9. In this study, the mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on photographs and tested on photographs were statistically

different at the $\alpha = 0.05$ level ($p = 0.038$). The mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on illustrations and tested on photographs were not statistically different ($p = 0.549$). The mean percentage accuracy of participants trained on photographs and tested on illustrations versus those of participants trained on photographs and tested on photographs were not statistically different ($p = 0.710$).

A.12 fMRI Study, Name-face Training without Feedback using Twelve Faces and Intermixed Blocks of Illustrations and Photographs

A.12.1 Participants

In this study, 12 University of Utah undergraduates participated. All participants were tested individually, none knew of the hypothesis being tested.

A.12.2 Materials

In this study, 12 face images from the AR Face Database [96], 6 males and 6 females, were used as stimulus. The images were presented to the participants as gray-scale photographs or as illustrations. An identical pilot study as in Study 1 was carried out and the facial illustrations were all rated as being a good likeness.

A.12.3 Procedure

This study was broken into two phases, a training phase and a testing phase. In both study and test phase faces are presented for 4.5 s, three faces per block, alternating the control block (repeated faces) and the test block (new faces). In the study phase, each face is presented twice. In the test phase, the first half presents the faces in the modality that they were learned. The second half presents the faces in the opposite modality. In the control periods, one male and one female face are repeated and the same name is presented on each side of the face. Order of periods

is illustration control, illustration test, photo control and photo test, repeated six times. In the study phase, one name is presented with the face, in the test phase, two different names are presented, and the participant picks the right or left name. The participants chose the name that they thought was correct by pressing a key on the Lumina keypad. During the testing phase, the participants were presented with each face with two names in the same modality, illustrations or photographs, with which the participant had been trained, for five seconds. The participants chose the name that they thought was correct by pressing a key on the laptop computer keyboard or pressed another key if they thought that this was a face that they had not previously seen. The modality, illustration or photograph, and the order of presentation were mixed between the training and testing phases of the study.

A.12.4   Design

1. Participants were given a consent form to read and sign.

2. Participants were then told that they would be participating in a study that would examine memory for faces.

3. The participants were read the following instructions. "I am going to introduce you to a group of people by presenting their names with their faces. Please try to remember the name of each person. You will have three study periods where you see each of the faces. After the study period, you will be presented with the faces with two different names, in same format as you saw them (photos) a different format (illustrations), or you will see new faces that you have not seen before. If the face is a new face, press the m key for new. If the face is one that you have seen before (either as a photo or a line drawing), choose the correct name by pressing either the z button for the left name, or the x button for the right name. First, we will show you all of the faces three times. You may rest between study sessions and press any key when you are ready to begin again. Do you have any questions?"

Condition	Min	Max	Mean	Std. Dev.
Train Illustration, Test Illustration	56	100	81.82	14.29
Train Illustration, Test Photograph	56	100	79.80	15.57
Train Photograph, Test Illustration	56	100	81.82	15.13
Train Photograph, Test Photograph	67	100	87.88	12.62

Table A.10. This table shows the minimum, maximum, and mean percentage of correctly identified faces for the four methods of training and testing.

4. Learning Session: Participants were then allowed to study thirty-two faces, shown to them as either illustrations or photographs, for five seconds each. Each face was shown three times.

5. Testing Session: Participants were presented with each face with two names for ten seconds. Participants chose the name they thought was correct using a key press on the laptop computer, or pressed another key to demonstrate that they believed that they had not seen the face before.

A.12.5 Results

Descriptive statistics for the rate of learning over the four learning and testing conditions are shown in Table A.10. In this study, the mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on photographs and tested on photographs were not statistically different ($p = 0.263$). The mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on illustrations and tested on photographs were not statistically different ($p = 0.754$). The mean percentage accuracy of participants trained on photographs and tested on illustrations versus those of participants trained on photographs and tested on photographs were not statistically different ($p = 0.399$).

A.13   Study 9: Illustrations versus Superportraits

A.13.1   Participants

In this study, 12 University of Utah undergraduates participated. All participants were tested individually, none knew of the hypothesis being tested.

A.13.2   Materials

In this study, 64 face images from the AR Face Database [96], 32 males and 32 females, were used as stimulus. Facial illustrations were created from the 64 images. The 64 facial illustrations, 32 males and 32 females, were all rated as being a good likeness by five independent judges. Using these 64 facial illustrations, 64 superportraits were created. First a male and a female norm face feature grid (FFG) were computed using the values of the FFGs for the male and female facial illustrations respectively. The facial feature grid is covered in Section 2.5. The difference between the norm face grid and the grid for each of the 64 facial illustrations was scaled by scaled by twenty percent and the source illustrations were warped correspondingly. Other than the difference in stimulus images the materials used in this study were the same as those reported in Appendix A.3.

A.13.3   Procedure

The procedure followed in this study was the same as that reported in Appendix A.4.

A.13.4   Design

The design of this study was the same as that reported in Appendix A.3, with superportraits used instead of photographs.

Condition	Min	Max	Mean	Std. Dev.
Train Illustration, Test Illustration	60	100	89.98	14.48
Train Illustration, Test Superportrait	50	100	81.98	16.28
Train Superportrait, Test Illustration	52	88	75.39	12.28
Train Superportrait, Test Superportrait	75	100	92.48	10.49

Table A.11. This table shows the minimum, maximum, and mean percentage of correctly identified faces for the four methods of training and testing. Participants were trained for fifteen seconds with each image.

A.13.5   Results

Descriptive statistics for the rate of learning over the four learning and testing conditions are shown in Table A.11. In this study, the mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on superportraits and tested on superportraits were not statistically different ($p = 0.505$). The mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on illustrations and tested on superportraits were not statistically different ($p = 0.267$). The mean percentage accuracy of participants trained on superportraits and tested on illustrations versus those of participants trained on superportraits and tested on superportraits were statistically different ($p < 0.001$). New superportraits were recognized as new faces 83 percent of the time and new illustrations were recognized as new faces 76 percent of the time. The mean accuracy for recognizing new faces from superportraits was not significantly different from the mean accuracy of recognizing new faces from illustrations ($p = 0.675$).

A.14   Study 10: Illustrations versus Caricatures

A.14.1   Participants

In this study, 12 University of Utah undergraduates participated. All participants were tested individually, none knew of the hypothesis being tested.

A.14.2 Materials

In this study, 64 face images from the AR Face Database [96], 32 males and 32 females, were used as stimulus. Facial illustrations were created from the 64 images. The 64 facial illustrations, 32 males and 32 females, were all rated as being a good likeness by five independent judges. Using these 64 facial illustrations, 64 caricatures were created. The caricatures were created using the free hand warping techniques described in Section 2.6 and were similar to those shown in Figure 2.7. Other than the difference in stimulus images the materials used in this study were the same as those reported in Appendix A.3.

A.14.3 Procedure

The procedure followed in this study was the same as that reported in Appendix A.4.

A.14.4 Design

The design of this study was the same as that reported in Appendix A.3, with caricatures used instead of photographs.

A.14.5 Results

Descriptive statistics for the rate of learning over the four learning and testing conditions are shown in Table A.12. In this study, the mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on caricatures and tested on caricatures were not statistically different ($p = 0.505$). The mean percentage accuracy of participants trained on illustrations and tested on illustrations versus those of participants trained on illustrations and tested on caricatures were not statistically different ($p = 0.267$). The mean percentage accuracy of participants trained on Caricatures and tested

Condition	Min	Max	Mean	Std. Dev.
Train Illustration, Test Illustration	50	100	86.46	15.50
Train Illustration, Test Caricature	50	100	77.98	18.48
Train Caricature, Test Illustration	38	88	73.81	13.39
Train Caricature, Test Caricature	75	100	90.48	9.49

Table A.12. This table shows the minimum, maximum, and mean percentage of correctly identified faces for the four methods of training and testing. Participants were trained with each image for fifteen seconds.

on illustrations versus those of participants trained on caricatures and tested on caricatures were statistically different ($p < 0.001$). New caricatures were recognized as new faces 76 percent of the time and new illustrations were recognized as new faces 83 percent of the time.

APPENDIX

ANALYZING FMRI STUDY DATA

This section discusses the analysis of raw data from an fMRI scan. The goal of the analysis is to determine those regions in the fMRI dataset in which signal changes occur upon stimulus presentation. The analysis of fMRI data falls into two parts. First the raw data are analyzed to produce an image showing regions of activation. Second, some level of significance must be calculated so that the probability of producing such a result purely by chance is suitably low [12]. This information can be used in turn to determine brain activation for a given stimulus. A volumetric dataset is represented as a three-dimensional discrete regular grid of volume elements (voxels) [77]. In the case of fMRI a voxel is a quantum unit of volume with an associated numeric value representing emitted energy.

The analysis of fMRI data can be broadly divided into three stages [44]:

1. Spatial Processing: including signal processing, data realignment, and smoothing.

2. Statistical Analysis.

3. Inference and Presentation.

Pre-processing steps are applied to the data to improve the detection of activation events. These include: registering the images to correct for participant movement during the study, and smoothing the data to improve the signal to noise ratio. Next, the statistical analysis is carried out to detect voxels in the dataset that show a response to stimulus. Finally, the activation images can be displayed. In addition, the statistical confidence placed in the result can be reported.

B.1 Spatial Processing

A number of pre-processing steps can be carried out on the raw fMRI dataset before statistical analysis of the data. Each of these pre-processing steps is independent of the others and each offers benefits as well as a computational time penalty that may be on the order of hours.

B.1.1 Raw Data Manipulation

The raw time data from the fMRI scanner requires Fourier transformation to form images. One type of artifact that can occur in fMRI is the Nyquist, or N/2 ghost [23, 24, 71]. This artifact is caused by acquiring odd or even echoes under the opposite gradient. In fMRI analysis, ghosting artifacts can cause a number of problems. Ghosting in activated regions could lead to apparent activation appearing outside the head. Effects that are more serious occur if the artifact changes with time. Movement of the participants' head causes the fringes of overlapping ghost and image areas to change dramatically. In this case, even small displacements may appear as large signal changes. Because participant motion is often correlated with stimulus, the changes of the interference pattern can show up in the statistical analysis as BOLD activation due to stimulus response. Correction for N/2 ghost artifact can be carried out on the raw time data by changing the phase angle between real and imaginary data points of alternate lines [24, 71]. The extent of correction required can be determined by adjusting the phase angle until the N/2 ghost in the first image in the set is minimized. The same correction is then applied to every image in the fMRI data set [24, 71].

Two data reducing steps can be performed that will cut down the number of calculations that need to be carried out in the analysis phase. Experimental pre-scans or saturation scans, which ensure that recovery effects have reached a steady state, can be removed from the beginning of the data set. In addition, the scan matrix size can be reduced so that it covers only the brain.

Changes in the global blood flow to the brain, as well as instability in the scanner hardware, can cause the mean intensity of the images to vary with time.

Such fluctuations mean that the response to each stimulus is not identical and reduces the power of any statistical test. In order to minimize this effect each image can be scaled so that it has an intensity mean equal to a pre-determined value such as the global intensity mean of the data set. Since there are fluctuations in image intensity caused by external interference, which are independent of the variation in the true image intensity, it is necessary to exclude these regions from the calculation of the image mean. Regions to exclude are chosen by the user before normalization.

B.1.2 Motion Correction

Changes in fMRI signal intensity over time occur due to participant head motion and can be confused with signal changes due to brain activity. Restrained and cooperative participants will still show displacement of up to a millimeter in the course of a study [48]. Therefore, the analysis of fMRI data starts with a series of spatial transformations to realign the data in order to undo the effects of participant movement during scanning. These transformations reduce variance in the time-series voxel data induced by movement or shape differences. In order to assign an observed response to a particular brain structure the fMRI data must conform to a known anatomical space. The fMRI data are therefore transformed using linear or nonlinear warps into a standard anatomical space in order to report in a frame of reference that can be related to other studies. Finally, the data are spatially smoothed before analysis in order to normalize the distribution of error in the fMRI dataset.

Realigning the volumetric fMRI dataset is a two-step process. First, each scan in the time series is co-registered to a target. Either the first scan in the time series or the average of all scans in the time series are generally used as the reference scan. In order to perform co-registration, a series of rigid-body transformations are estimated by minimizing a function of the difference between the current and reference scans. The six rigid-body transformations are translations in the x, y, and z directions and rotations about the x, y, and z-axes. A least squares solution

yielding estimations of the six rigid-body transformation parameters is calculated for each scan. The rigid-body transformations are computed using a first order approximation of the Taylor expansion of the effect of movement on signal intensity using spatial derivatives of slices of the fMRI dataset. This technique allows for an iterative least squares solution corresponding to a Gauss-Newton search [45, 145].

In the second part of the realignment process the rigid-body transformations are applied to the fMRI dataset by re-sampling the dataset according to the spatial transformation estimated in the first stage of the realignment process. The numeric value corresponding to fMRI signal intensity of each voxel in the transformed dataset is determined from the intensity of surrounding voxels in the original dataset. The simplest method of image re-sampling is nearest neighbor re-sampling. In nearest neighbor re-sampling the value of the closest voxel is taken as the value of the transformed voxel. This approach has the advantage of preserving the original intensities, but can severely degrade the dataset. In order to realign an fMRI dataset with subvoxel accuracy, the spatial transformations must involve fractions of a voxel. It is therefore necessary to re-sample the fMRI dataset at positions between the centers of voxels. This requires an interpolation scheme to estimate the intensity of a voxel, based on the intensity of its neighbors [57]. The rigid-body transformations are applied by re-sampling the fMRI data using trilinear, sinc, or cubic spline interpolation.

Trilinear interpolation is the simplest form of interpolation for three-dimensional data. The interpolated value is computed by a linear combination of the values of the eight neighboring voxels [66]. However, Trilinear interpolation can introduce sampling errors because some high frequency information can be removed from the volume dataset [106].

The ideal scheme for transforming band-limited MR images without introducing artifacts is to perform the translations and rotations in Fourier space. This approach is referred to as Fourier interpolation and has been implemented in two dimensions [41]. The image space method that gives the closest results to Fourier interpolation is a full sinc interpolation using every voxel in the image to calculate

the new value at a single voxel [73]. Although sinc interpolation is the theoretically optimal convolution kernel for band-limited images, it may not be the ideal kernel in terms of application. The fMRI volumetric data have finite spatial support, which implies that it is not possible for the sampling frequencies to satisfy the Nyquist criterion. Consequently, it is impossible to retrieve the original fMRI data exactly from the resulting samples by means of sinc interpolation [98]. Another problem with sinc interpolation is that the sinc function has infinite support. Therefore, it cannot be computed in practice for most medical imaging applications [98]. Accurate realignment of the fMRI volumetric data is the slowest procedure in fMRI data analysis [57]. Because of the computational burden of performing a full sinc interpolation, in practice it is necessary to limit the extent of the sinc function [59]. The truncated sinc function is commonly implemented using an 11-voxel Hamming window sinc function [57].

The results of a number of studies show that spline interpolation constitutes the best trade-off between accuracy and computational cost in medical imaging applications [85, 86, 98, 106, 135]. Splines constitute an elegant framework for dealing with interpolation and discretization problems [135–138]. Splines are widely used in computer-aided design and computer graphics, but have been until recently neglected in medical imaging. Most forms of spline fitting (interpolation, least squares approximation, smoothing splines) can be performed efficiently using recursive digital filters [135]. In addition, the multi-resolution properties of splines make them amenable to constructing wavelet bases and computing image pyramids both of which have medical imaging applications.

Realignment algorithms based on the co-registration approach can achieve accuracy in the range between fifty and one hundred microns [48, 50, 76]. In fMRI however, there may be additional, movement related, nonlinear sources of error. Friston et al. show, in extreme cases, up to ninety percent of the variance in an fMRI dataset can be accounted for by the effects of movement after realignment [48]. These residual errors are due to movement effects that cannot be modeled with a linear affine model. These nonlinear effects include: participant movement between

slice acquisition, interpolation artifacts [57], nonlinear distortion due to magnetic field inhomogeneities [5] and spin excitation history effects [48].

B.1.3 Smoothing

Any reduction in the random noise in the data improves ability of a statistical technique to detect activation [104]. Therefore, spatially and temporally smoothing the fMRI data improves the validity of statistical inferences by increasing the signal-to-noise ratio. However, smoothing will reduce the resolution in each image, therefore a balance must be achieved between improving the signal-to-noise ratio and maintaining functional resolution in the fMRI data. There is no straightforward answer to the question of which is the best smoothing width to use in the analysis of an fMRI data set.

Improvements in the signal-to-noise ratio can be made by smoothing in the temporal domain as well as in the spatial domain. The BOLD contrast effect and the rate at which the fMRI signal changes in an active brain region are limited by blood flow. Hence, temporal smoothing with a filter, which approximates the hemodynamic response function, will improve the signal-to-noise ratio of the fMRI dataset. A three-dimensional Gaussian filter of width 2.8 seconds is reported to be a good approximation to the hemodynamic response function [46].

B.2 Statistical Analysis

The most straightforward way to analyze fMRI data is to subtract the mean *off* data from the mean *on* data. The disadvantage of this brute force subtraction technique is that small movements of the participants' head can drastically change the intensity of voxels at the boundaries of the dataset. Such an artifact can give rise to a ring of apparent activation near the brain boundaries. A t-test can be used to reduce the activation ring effect and to yield a statistic that can be tested against the null hypothesis that no signal is present. The t-test is used to reduce movement artifacts by giving high t-scores to large differences with small standard deviations, and low t-scores to small differences with large standard deviations. An

image or dataset where values are assigned based on the output of a statistical test is commonly called a statistical parametric map.

Statistical parametric mapping (SPM) is used to identify functionally specialized brain regions and is the most prevalent approach in characterizing functional brain anatomy. SPM is a voxel-based approach, which allows an analyst to make comments about brain response to a given stimulus. SPM can be made with different experimental designs. However, fMRI data lends itself to a signal processing perspective [44].

A commonly used technique is correlation coefficient mapping. In this technique, the time response of the activation to the stimulus is predicted, using prior knowledge of the hemodynamic response, and a correlation coefficient for each voxel over time is calculated [47]. The major disadvantage of this technique is that it is particularly sensitive to motion artifacts. Other methods that have been used include Fourier transformation, (identifies pixels with a high Fourier component [84]), principal component analysis, (locates regions in the brain that show synchronous activity using eigenfunctions [11]), clustering techniques, (compute synchrony using iterative methods [43]), and various non-parametric tests that do not require the assumption of normality in the signal distribution [67]. All these have strengths and weaknesses. The necessary criteria for a technique to be successful are sensitivity, simplicity, speed, and statistical validity.

B.3   Inference and Presentation

After computing a statistical map, it is necessary to display the regions of activation and an estimate of the reliability of the result. If the distribution of the statistic, under the null hypothesis of no activation, is known then statistical tables can be used to threshold the image. After thresholding, only those pixels with a strong stimulus correlation are shown.

After thresholding to display only the active brain regions for a given stimulus, the active regions are superimposed on background images to enable anatomical localization. If the fMRI data contains a high-resolution component, this can be

used for base images. Better contrast may come from some inversion recovery images that show only white matter or gray matter. Figures B.1, B.2, B.3 and B.4 on the following four pages show examples of the active brain regions from the fMRI study superimposed on a high-resolution average brain image. These images are of the same data shown in Figure 3.11 of Chapter 3.
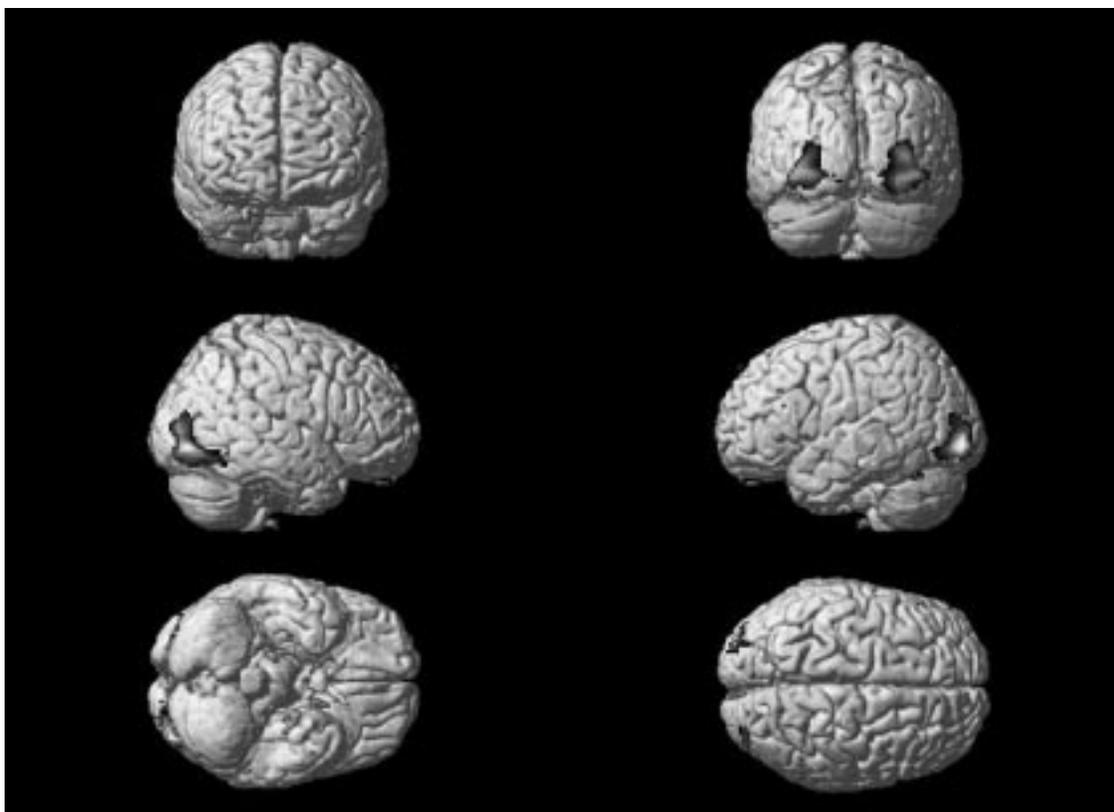
Figure B.1. These images show active brain regions from the fMRI study during the training phase using photographs as stimulus superimposed on a high-resolution average brain image. These images show thresholded activation during the training phase of the fMRI study with photographs as the stimulus (encoding photographs). These images are computed over all of the participants in the fMRI study. These images can be compared to the image in the upper left of Figure 3.11 of Chapter 3
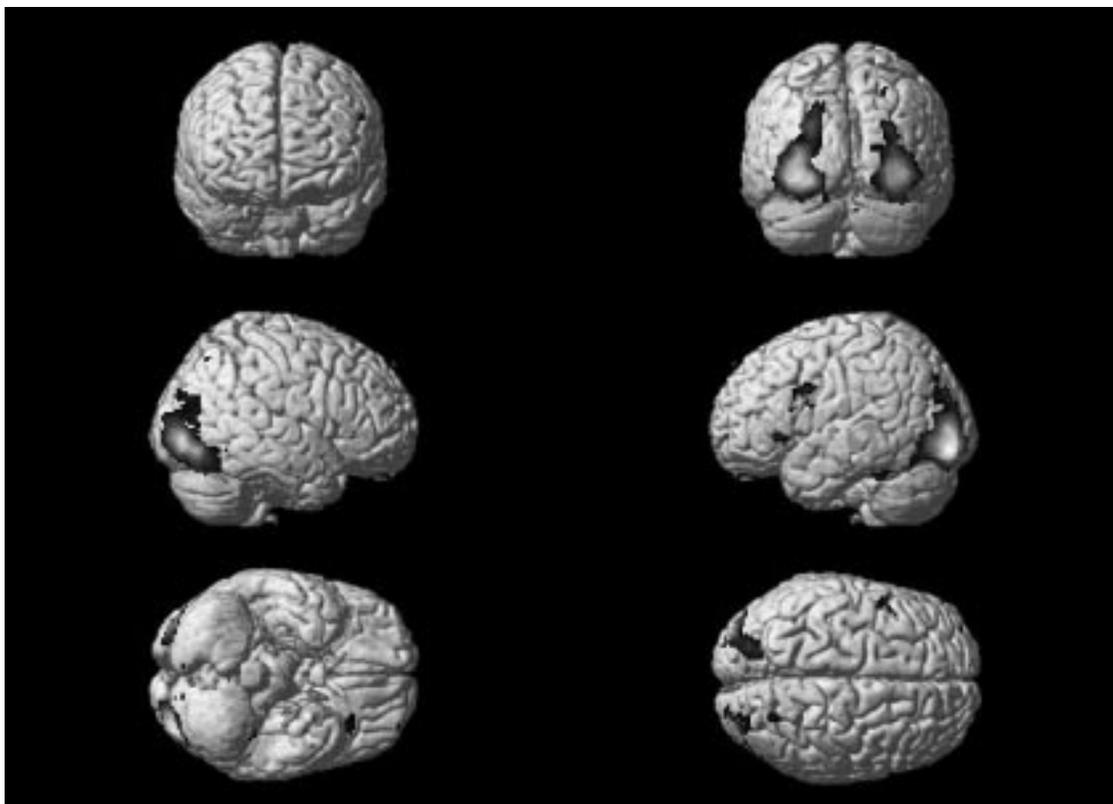
Figure B.2. These images show active brain regions from the fMRI study during the testing phase using photographs as stimulus superimposed on a high-resolution average brain image. These images show thresholded activation during the testing phase of the fMRI study with photographs as the stimulus (decoding photographs). These images are computed over all of the participants in the fMRI study. These images can be compared to the image in the upper right of Figure 3.11 of Chapter 3.
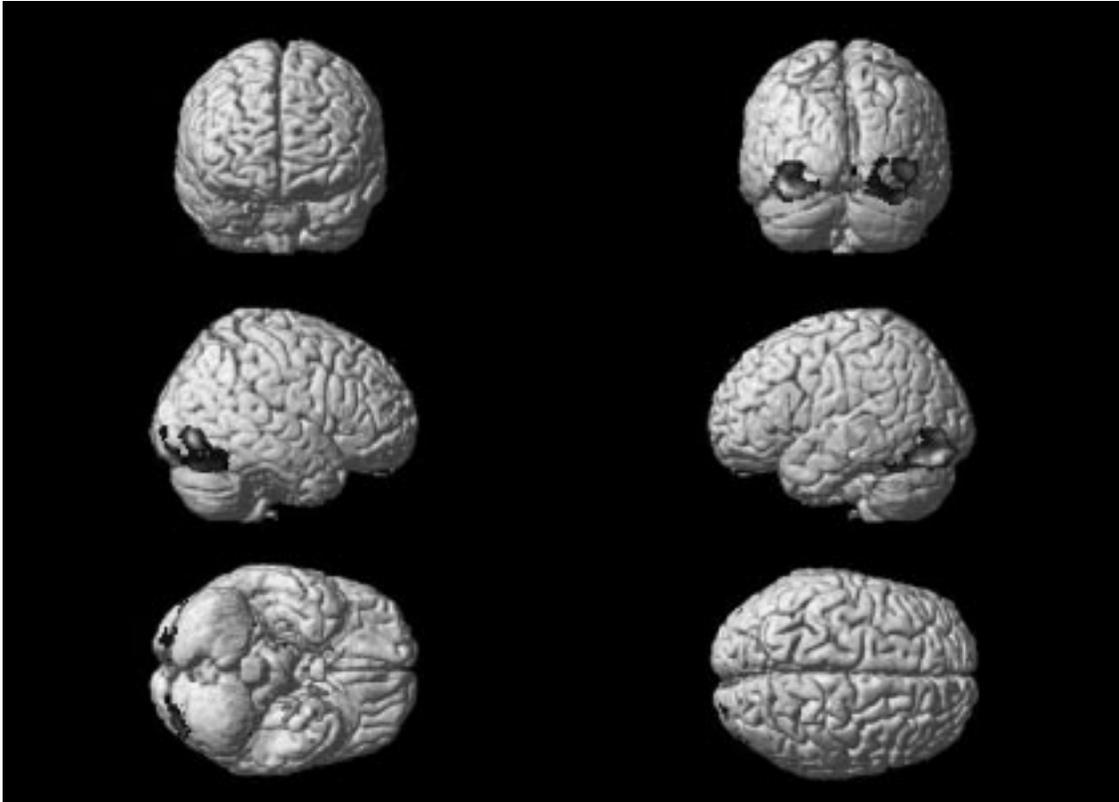
Figure B.3. These images show active brain regions from the fMRI study during the training phase using illustrations as stimulus superimposed on a high-resolution average brain image. These images show thresholded activation during the training phase of the fMRI study with illustrations as the stimulus (encoding illustrations). These images are computed over all of the participants in the fMRI study. These images can be compared to the image in the lower left of Figure 3.11 of Chapter 3.
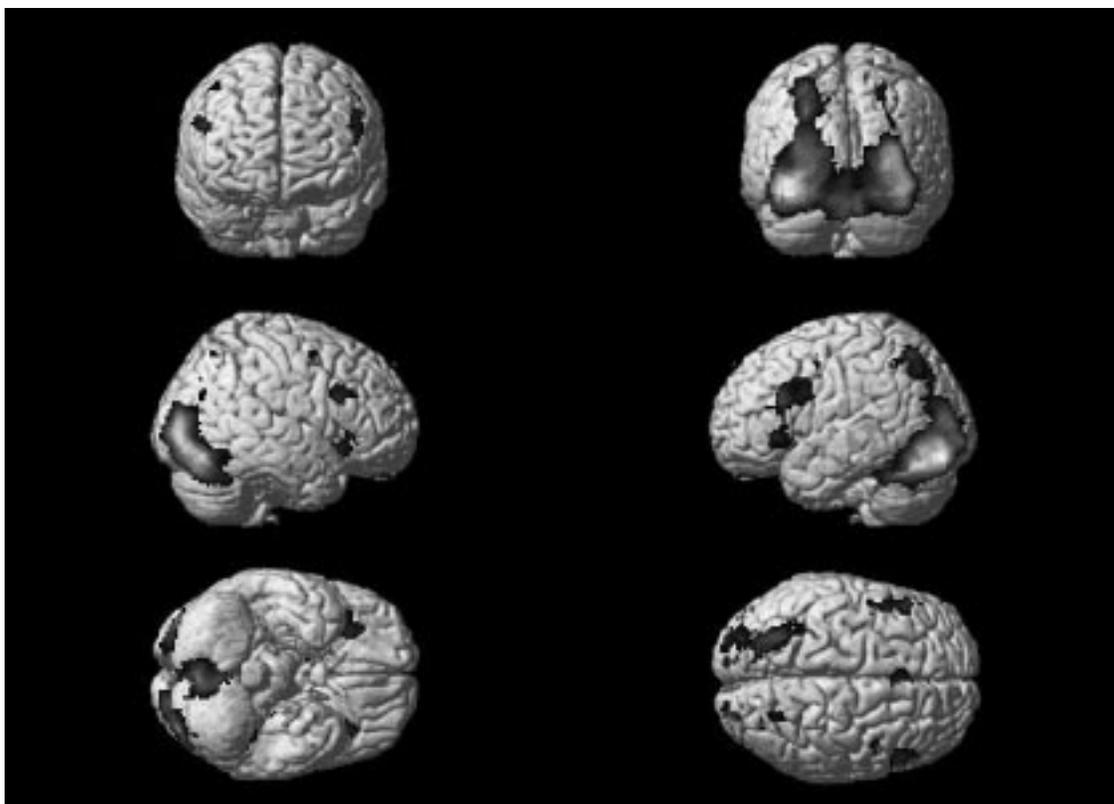
Figure B.4. These images show active brain regions from the fMRI study during the testing phase using illustrations as stimulus superimposed on a high resolution average brain image. These images show thresholded activation during the testing phase of the fMRI study with illustrations as the stimulus (decoding illustrations). These images are computed over all of the participants in the fMRI study. These images can be compared to the image in the lower right of Figure 3.11 of Chapter 3.

REFERENCES

[1] ABBOUD, H. A. Cedrus corporation, http://www.cedrus.com.

[2] ADELSON, E., AND BURT, P. Image data compression with the laplacian pyramid, 1981.

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

[4] ALLEBACH, J. Visual model-based algorithms for halftoning images, 1981.

[5] ANDERSSON, J. L., HUTTON, C., ASHBURNER, J., TURNER, R., AND FRISTON, K. Modeling geometric deformations in epi time series. *NeuroImage 13* (2001), 903–919.

[6] ANDREWS, B. A., AND POLLEN, D. A. Relationship between spatial frequency selectivity and receptive field profile of simple cells. *Journal of Physiology 287* (1979), 163–176.

[7] AREND, L., AND GOLDSTEIN, R. Lightness models, gradient illusions, and curl. *Perception and Psychophysics 43* (1987), 65–80.

[8] ARTHURS, O. J., AND BONIFACE, S. How well do we understand the neural origins of the fmri bold signal? *Trends in Neurosciences 25*, 1 (2002), 27–31.

[9] ATICK, J. J., AND REDLICH, N. A. What does the retina know about natural scenes? *Neural Computation 4* (1992), 196–210.

[10] AVIDAN, G., AND BEHRMANN, M. Correlations between the fmri bold signal and visual perception. *Neuron 34*, 4 (2002), 495–497.

[11] BACKFRIEDER, W., BAUMGARTNER, R., SMAL, M., MOSER, E., AND BERGMANN, H. Quantification of intensity variations in functional mr images using rotated principal components. *Physical Medical Biology 41* (1996), 1425–1438.

[12] BANDETTINI, P. A., JESMANOWICZ, A., WONG, E. C., AND HYDE, J. S. Processing strategies for time-course data sets in functional mri of the human brain. *Magnetic Resonance Medicine 30* (1993), 161–173.

[13] BANDETTINI, P. A., WONG, E. C., JESMANOWICZ, A., , HINKS, R. S., AND HYDE, J. S. Spin-echo and gradient-echo epi of human brain activation using bold contrast: a comparative study at 1.5t. *NMR Biomed 7* (1992), 12–20.

[14] BARCH, D. M., SABB, F. W., CARTER, C. S., BRAVER, T. S., NOLL, D. C., AND COHEN, J. D. Overt verbal responding during fmri scanning: Empirical investigations of problems and potential solution. *NeuroImage 10* (1999), 642–657.

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

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

[17] BIEDERMAN, I., AND KALOCSAI, P. Neural and psuchological analysis of object and face recognition. In *Face Recognition: From Theory to Applications* (1998), Springer-Verlag, pp. 3–25.

[18] BISHOP, P. O., COOMBS, J. S., AND HENRY, G. H. Receptive fields of simple cells in the cat striate cortex. *Journal of Physiology 231* (1973), 31–60.

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

[20] BOFF, K., KAUFMAN, L., AND THOMAS, J. P. *Handbook of Perception and Human Performance.* John Wiley & Sons, 1986.

[21] BOYNTON, G. M., ENGEL, S. A., GLOVER, G. H., AND HEEGER, G. H. Linear systems analysis of functional magnetic resonance imaging in human v1. *The Journal of Neuroscience 16* (1996), 13.

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

[23] BRUDER, H., FISCHER, H., REINFELDER, H. E., , AND SCHMITT, F. Image reconstruction for echo planar imaging with nonequidistant k-space sampling. *Magnetic Resonance Medicine 23* (1992), 311–323.

[24] BUONOCORE, M. H., AND GAO, L. Ghost artifact reduction for echo planar imaging using image phase correction. *Magnetic Resonance Medicine 38* (1997), 89–100.

[25] BURT, P. J., AND ADELSON, E. H. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications COM-31,4* (1983), 532–540.

[26] BURT, P. J., AND ADELSON, E. H. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics 2*, 4 (1983), 217–236.

[27] BURTON, A. M., BRUCE, V., AND HANCOCK, P. J. B. From pixels to people: a model of familiar face recognition. *Cognitive Science 23*, 1 (1999), 1–31.

[28] Buxton, R., and Frank, L. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *jcbfm 17*, 1 (1997), 64–72.

[29] Canny, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 8* (1986), 679–698.

[30] Canny, J. F. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 8* (1986), 769–798.

[31] Cherished Moments. http://www.cherishedmoments.com/most-popular-baby-names.htm.

[32] Ciuciu, P., Poline, J. B., Marrelec, G., Idier, J., Pallier, C., and Benali, H. Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. Tech. rep., SHFJ/CEA, Mar. 2002.

[33] Cohen, M. A., and Grossberg, S. Neural dynamics of brightness perception: features, boundaries, diffusion and resonance. *Perception and Psychophysics 36*, 5 (1984), 428–456.

[34] Cosman, P. C., Gray, R. M., and Vetterli, M. Vector quantization of image subbands: A survey. *IEEE Transactions on Image Processing 5*, 2 (1996), 202–225.

[35] Dale, A. M. Optimal experimental design for event-related fmri. *Human Brain Mapping 8* (1999), 109–114.

[36] Daugman, J. G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Optical Society of America 2*, 7 (1985), 1160–1169.

[37] Davies, G., Ellis, H., and Shepperd, J. Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology 63*, 2 (1978), 180–187.

[38] Denecker, K., Ville, D. V. D., Habils, F., Meeus, W., Brunfaut, M., and Lemahieu, I. Design of an improved lossless halftone image compression codec. *Signal Processing: Image Communication 17* (2002), 277–292.

[39] Durand, F., Ostromoukhov, V., Miller, M., Duranleau, F., and Dorsey, J. Decoupling strokes and high-level attributes for interactive traditional drawing. In *Rendering Techniques 2001: 12th Eurographics Workshop on Rendering* (June 2001), Eurographics, pp. 71–82. ISBN 3-211-83709-4.

[40] Eddy, W. F., Fitzgerald, M., Genovese, C., Lazar, N., Mockus, A., and Welling, J. The challenge of functional magnetic resonance imaging. *Journal of Computational and Graphical Statistics 8*, 3 (1999), 545–??.

[41] EDDY, W. F., FITZGERALD, M., AND NOLL, D. C. Improved image registration by using fourier interpolation. *Magnetic Resonance Medicine 36* (1996), 923–931.

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

[43] FORMAN, S. D., COHEN, J. D., EDDY, M. F. N. W. F., MINTUN, M. A., AND NOLL, D. C. Improved assessment of significant activation in functional magnetic resonance imaging (fmri): Use of a cluster-size threshold. *Magnetic Resonance Medicine 33* (1995), 636–647.

[44] FRISTON, K. J. *Statistical Parametric Mapping.* The Wellcome Trust, 2000.

[45] FRISTON, K. J., ASHBURNER, J., FRITH, C. D., POLINE, J.-B., HEATHER, J. D., AND FRACKOWIAK, R. S. J. Spatial registration and normalization of images. *Human Brain Mapping 2* (1995), 165–189.

[46] FRISTON, K. J., HOLMES, A. P., POLINE, J.-B., GRASBY, P. J., WILLIAMS, S. C. R., FRACKOWIAK, R. S. J., AND TURNER, R. Analysis of fmri time-series revisited. *Neuroimage 2* (1995), 45–53.

[47] FRISTON, K. J., JEZZARD, P., AND TURNER, R. Analysis of functional mri time series. *Human Brain Mapping 1* (1994), 153–171.

[48] FRISTON, K. J., WILLIAMS, S., HOWARD, R., FRACKOWIAK, R. S. J., AND TURNER, R. Movement related effects in fmri time series. *Magnetic Resonance Medicine 35* (1996), 346–355.

[49] FRISTON, K. J., WORSLEY, K. J., FRACKOWIAK, R. S. J., MAZZIOTTA, J. C., AND EVANS, A. C. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping 1* (1994), 214–220.

[50] FROUIN, V., MESSEGUE, E., AND MANGIN, J.-F. Assessment of two fmri motion correction algorithms. *Human Brain Mapping 5* (1997), 485.

[51] GABOR, D. Theory of communication. *Journal of IEEE vol. 93* (1946), 429–457.

[52] GADIAN, D. G. *NMR and its applications to living systems.* Oxford university Press, 1995.

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

[54] GILCHRIST, A. *Lightness Brightness and Transparency.* Lawrence Erlbaum Associates, 1994.

[55] GOVE, A., GROSSBERG, S., AND MINGOLLA, E. Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience 12* (1995), 1027–1052.

[56] GRAY, R. M., PERLMUTTER, K., AND OLSHEN, R. A. Quantization, classification, and density estimation for kohonen's gaussian mixture. In *Data Compression Conference* (1998), pp. 63–72.

[57] GROOTOONK, S., HUTTON, C., ASHBURNER, J., HOWSEMAN, A. M., JOSEPHS, O., REES, G., FRISTON, K. J., AND TURNER, R. Characterization and correction of interpolation effects in the realignment of fmri time series. *NeuroImage 11* (2000), 49–57.

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

[59] HAJNAL, J. V., SAEED, N., SOAR, E. J., OATRIDGE, A., YOUNG, I. R., AND BYDDER, G. M. A registration and interpolation procedure for subvoxel matching of serially acquired mr images. *J. Comput. Assist. Tomogr. 19* (1995), 289–296.

[60] HALL, D. A., HAGGARD, M. P., AKEROYD, M. A., SUMMERFIELD, A. Q., PALMER, A. R., ELLIOTT, M. R., AND BOWTELL, R. W. Modulation and task effects in auditory processing measured using fmri. *Human Brain Mapping 10* (2000), 107–119.

[61] HALL, D. A., SUMMERFIELD, A. Q., GONALVES, M. S., FOSTER, J. R., PALMER, A. R., AND BOWTELL, R. W. Time-course of the auditory bold response to scanner noise. *Magnetic Resonance Medicine 43* (2000), 601–606.

[62] HANSEN, M. H. Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology 58* (1959), 3221–3334.

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

[64] HATHOUT, G. M., GOPI, R. K., BANDETTINI, P., AND GAMBHIR, S. S. The lag of cerebral hemodynamics with rapidly alternating periodic stimulation: Modeling for functional mri. *Magnetic Resonance Imaging 17* (1999), 9–20.

[65] HEEGER, D. J., AND RESS, D. What does fMRI tell us about neuronal activity? *Nature Reviews Neuroscience 3* (2002), 142–151.

[66] HILL, S. *Tri-linear interpolation*. Academic Press Professional, Inc., 1994.

[67] HOLMES, A. P., BLAIR, R. C., WATSON, J. D. G., AND FORD, I. Nonparametric analysis of statistic images from functional mapping experiments. *Blood Flow Metabolism 16* (1996), 7–22.

[68] HORN, B. K. P. Determining lightness from an image. *CVGIP 3* (1974), 277–299.

[69] HOWSEMAN, A. M., AND BOWTELL, R. W. Functional magnetic resonance imaging: imaging techniques and contrast mechanisms. *Philosophical Transactions of the Royal Society 354* (1999), 1179–1194.

[70] HOWSEMAN, A. M., PORTER, D. A., HUTTON, C., JOSEPHS, O., AND TURNER, R. Blood oxygenation level dependent signal time courses during prolonged visual stimulation. *Magnetic Resonance Imaging 16* (1998), 1–11.

[71] HU, X., AND TUONG, H. L. Artifact reduction in epi with phase-encoded reference scan. *Magnetic Resonance Medicine 36* (1996), 166–171.

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

[73] JAIN, A. K. *Fundamentals of Digital Image Processing.* Prentice Hall, 1989.

[74] JERNIGAN, M. E., AND MCLEAN, G. F. 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, 1992, ch. 17, pp. 451–462.

[75] JEZZARD, P., HEINEMANN, F., TAYLOR, J., DESPRES, D., WEN, H., BALABAN, R. S., AND TURNER, R. Comparison of epi gradient-echo contrast changes in cat brain caused by respiratory challenges with direct simultaneous evaluation of cerebral oxygenation via a cranial window. *NMR in Biomedicine 7* (1994), 35–44.

[76] JIANG, A. P., KENNEDY, D. N., BAKER, J. R., WEISSKOFF, R. M., TOOTELL, R. B. H., WOODS, R. P., BENSON, R. R., KWONG, K. K., BRADY, T. J., ROSEN, B. R., AND BELLIVEAU, J. W. Motion detection and correction in functional mr imaging. *Human Brain Mapping 3* (1995), 224–235.

[77] KAUFMAN, A. Volume visualization. *IEEE Computer Society Press Tutorial* (1991).

[78] KENNEDY, J. M. *A psychology of picture perception.* Jossey-Bass Publishers, San Fransisco, 1974.

[79] KNUTH, D. E. Digital halftones by dot diffusion. *ACM Transactions on Graphics (TOG) 6*, 4 (1987), 245–273.

[80] KULIKOWSKI, J. J., MARCELJA, S., AND BISHOP, P. O. Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biological Cybernetics 43* (1982), 187–198.

[81] KWONG, K. K. Functional magnetic resonance imaging with echo planar imaging. *Magnetic Resonance Quarterly 11* (1995), 1–20.

[82] KWONG, K. K., BELLIVEAU, J. W., CHESLER, D. A., GOLDBERG, I. E., WEISSKOFF, R. M., AND ET AL. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Procedings of the National Acadamy of Science 89* (1992), 5675–5679.

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

[84] LANGE, N., AND ZEGER, S. L. Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Magnetic Resonance Medicine 46* (1997), 1–29.

[85] LEHMANN, T. M., GNNER, C., AND SPITZER, K. Survey: Interpolation methods in medical image processing. *IEEE Transactions on Medical Imaging 18*, 11 (1999), 1049–1075.

[86] LEHMANN, T. M., GNNER, C., AND SPITZER, K. Addendum: B-spline interpolation in medical image processing. *IEEE Transactions on Medical Imaging 20*, 7 (2001), 660–665.

[87] LEWIS, M. B., AND JOHNSTON, R. A. Understanding caricatures of faces. *The Quarterly Journal of Experimental Psychology 50A*, 2 (1998), 321–346.

[88] LOGOTHETIS, N. K., PAULS, J., AUGATH, M., TRINATH, T., AND OELTERMANN, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature 412*, 6843 (July 2001), 150–157.

[89] MACHIN, R. http://www.ricmachin.com/.

[90] MAO, H., AND BERNS, G. S. Mri in the study of brain functions: clinical perspectives. *Medicamundi 46* (2002), 1.

[91] MARCELJA, S. Mathematical description of the responses of simple cortical cells. *Optical Society of America 70*, 11 (1980), 1297–1300.

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

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

[94] MARRELEC, G., BENALI, H., CIUCIU, P., AND POLINE, J. B. Bayesian estimation of the hemodynamic response function in functional MRI. In *Bayesian Inference and Maximum Entropy Methods* (Aug. 2001), R. Fry, Ed., MaxEnt Workshops.

[95] MARTINEZ, A. http://rvl1.ecn.purdue.edu/ aleix/aleix_face_db.html.

[96] MARTINEZ, A., AND BENAVENTE, R. The ar face database. *CVC Technical Report 24* (1998).

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

[98] MEIJERING, E. H. W., NIESSEN, W. J., AND VIERGEVER, M. A. Quantitative evaluation of convolution-based methods for medical image interpolation. *Medical Image Analysis 5*, 2 (2001), 111–126.

[99] NOLL, D. C. A primer on mri and functional mri.

[100] OGAWA, S., AND LEE, T. M. Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurments and image simulation. *Magnetic Resonance Medicine 16* (1990), 9–18.

[101] OGAWA, S., LEE, T. M., NAYAK, A. S., AND GLYNN, P. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance Medicine 14* (1990), 68–78.

[102] OGAWA, S., LEE, T. M., AND TANK, D. W. Magnetic resonance imaging with contrast dependent on blood oxygenation. *Procdings of the National Acadamy of Science 87* (1990), 9868–9872.

[103] OGAWA, S., TANK, D. W., MENON, R., ELLERMANN, J. M., KIM, S. G., AND MERKLE, H. Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Procdings of the National Acadamy of Science 89* (1992), 5951–5955.

[104] OPPENHEIM, A. V. *Applications of Digital Signal Processing.* Prentice Hall, 1978.

[105] OSTROMOUKHOV, V. Digital facial engraving. *Proceedings of SIGGRAPH 99* (1999), 417–424.

[106] OSTUNI, J. L., SANTHA, A. K. S., MATTAY, V. S., WEINBERGER, D. R., LEVIN, R. L., AND FRANK, J. A. Analysis of interpolation effects in the reslicing of functional mr images. *Journal of Computer Assisted Tomography 21* (1997), 803–810.

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

[108] PAULING, L., AND CORYELL, C. D. The magnetic properties and structure of haemoglobin, oxyhaemoglobin and carbon monoxyhaemoglobin. *Procdings of the National Acadamy of Science* (1936).

[109] PEARSON, D. E., HANNA, E., AND MARTINEZ, K. Computer-generated cartoons. In *Images and understanding* (Cambridge, 1990), Cambridge University Press.

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

[111] PESSOA, L., MINGOLLA, E., AND NEUMANN, H. A contrast- and luminance-driven multiscale network model of brightness perception. *Vision Research 35*, 15 (1995), 2201–2223.

[112] POLLEN, D. A., LEE, J. R., AND TAYLOR, J. H. How does the striate cortex begin the reconstruction of the visual world? *Science 173* (1971), 74–77.

[113] POLLEN, D. A., AND RONNER, S. Phase relationships between adjacent cells in the visual cortex. *Science 212*, 4 (1981), 1409–1411.

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

[115] REDMAN, L. How to draw caricatures. In *Aspects of Face Processing* (Chicago Illinois, 1984), Contemporary Books.

[116] REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. Photographic tone reproduction for digital images. *ACM Transactions on Graphics, Proceedings Siggraph 2002* (2002). http://www.cs.ucf.edu/˜ reinhard/cdrom/.

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

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

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

[120] ROLLS, E. T. Functions of the primate temporal review lobe cortical visual areas in invariant visual object and face recognition. *Neuron 27* (2000), 205–218.

[121] ROSEN, B. R., BUCKNERAND, R. L., AND DALE, A. M. Event-related functional MRI: past, present and future. *Procedings of the National Acadamy of Science 95* (1998), 773–780.

[122] SABATINI, S. P. Recurrent inhibition and clustered connectivity as a basis for gabor-like receptive fields in the visual cortex. *Biological Cybernetics 74* (1996), 189–202.

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

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

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

[126] SINGH, K. D., BARNES, G. R., HILLEBRAND, A., FORDE, E. M. E., AND WILLIAMS, A. L. Task-related changes in cortical synchronization are spatially coincident with the hemodynamic response. *NeuroImage 16* (2002), 103–114.

[127] SOUSA, M. C., AND BUCHANAN, J. W. Observational model of blenders and erasers in computer-generated pencil rendering. In *Graphics Interface '99* (June 1999), pp. 157–166. ISBN 1-55860-632-7.

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

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

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

[131] THULBORN, K. R., WATERTON, J. C., MATTEWS, P. M., AND RADDA, G. K. Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochem Biophys Acta* (1982), 265–270.

[132] TURNER, R., BIHAN, D. L., MOONEN, C. T. W., DESPRES, D., AND FRANK, J. Echo-planar time course mri of cat brain oxygenation changes. *Magnetic Resonance Medicine 22* (1991), 159–166.

[133] TURNER, R., HOWSEMAN, A., REES, G. E., JOSEPHS, O., AND FRISTON, K. Functional magnetic resonance imaging of the human brain: data acquisition and analysis.

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

[135] UNSER, M. Splines: A perfect fit for signal and image processing. *IEEE Signal Processing Magazine 16*, 6 (1999), 22–38.

[136] UNSER, M., ALDROUBI, A., AND EDEN, M. Fast b-spline transforms for continuous image representation and interpolation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 13*, 3 (1991), 277–285.

[137] UNSER, M., ALDROUBI, A., AND EDEN, M. B-spline signal processing: Part i theory. *IEEE Transactions on Signal Processing 41*, 2 (1993), 821–833.

[138] UNSER, M., ALDROUBI, A., AND EDEN, M. B-spline signal processing: Part ii efficient design and applications. *IEEE Transactions on Signal Processing 41*, 2 (1993), 834–848.

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

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

[141] WANDELL, B. A. Computational neuroimaging of human visual cortex. *Annual Review Neuroscience 22* (1999), 145–173.

[142] WATANABE, T., HARNER, A. M., MIYAUCHI, S., SASAKI, Y., NIELSEN, M., PALOMO, D., AND MUKAI, I. Task-dependent influences of attention on the activation of human primary visual cortex. *Procedings of the National Acadamy of Science 95* (1998), 11489–11492.

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

[144] WONG, E. Artistic rendering of portrait photographs. Master's thesis, Cornell University, 1999.

[145] WOODS, R. P., CHERRY, S. R., AND MAZZIOTTA, J. C. Rapid automated algorithm for aligning and reslicing pet images. *Journal of Computer Assisted Tomography 16* (1992), 620–633.

[146] WORSLEY, K. J., LIAO, C. H., ASTON, J., PETRE, V., DUNCAN, G. H., MORALES, F., AND EVANS, A. C. A general statistical analysis for fMRI data. Tech. rep., Dept. of Maths. and Stats., Mc Gill University, 2001.

[147] YARMEY, A. D. I recognize your face but I can't remember your name: Further evidence on the tip-of-the-tongue phenomenon. *Memory and Cognition 1* (1973), 287–290.

[148] YOUNG, A. W., HAY, D. C., AND ELLIS, A. W. The face that launched a thousand slips: Everyday difficulties and errors in recognizing people. *British Journal of Psychology 76* (1985), 495–523.