TOWARDS THE DEVELOPMENT OF TRAINING TOOLS

FOR FACE RECOGNITION

A Dissertation

by

JOBANY RODRIGUEZ

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2011

Major Subject: Computer Engineering

TOWARDS THE DEVELOPMENT OF TRAINING TOOLS

FOR FACE RECOGNITION

A Dissertation

by

JOBANY RODRIGUEZ

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Ricardo Gutierrez-Osuna
Committee Members,	Yoonsuck Choe
	Dezhen Song
	Takashi Yamauchi
Head of Department,	Valerie E. Taylor

May 2011

Major Subject: Computer Engineering

ABSTRACT

Towards the Development of Training Tools

for Face Recognition. (May 2011)

Jobany Rodríguez, B.S., University of Puerto Rico at Mayagüez;

M.S., Monmouth University

Chair of Advisory Committee: Dr. Ricardo Gutierrez-Osuna

Distinctiveness plays an important role in face recognition, i.e., a distinctive face is usually easier to remember than a typical face. This distinctiveness effect explains why caricatures are recognized faster and more accurately than unexaggerated (i.e., veridical) faces. Furthermore, using caricatures during training can facilitate recognition of a person's face at a later time. The objective of this dissertation is to determine the extent to which photorealistic computer-generated caricatures may be used in training tools to improve recognition of faces by humans. To pursue this objective, we developed a caricaturization procedure for three-dimensional (3D) face models, and characterized face recognition performance through a series of perceptual studies.

The first study tested whether exposure to caricatures during an initial familiarization phase would aid in the recognition of their veridical counterparts at a

iv

later time. Results indicate that people are more accurate at recognizing unaltered faces if they are first familiarized with caricatures of the faces, rather than with the unaltered faces. In the second study, we sought to determine the extent to which familiarization with caricaturized faces could also be used to reduce other-race effects. Caucasian participants were first familiarized with faces from multiple races, and then asked to recognize those faces among a set of confounders. Participants who were familiarized with and then asked to recognize veridical versions of the faces showed a significant other-race effect on Indian faces. In contrast, participants who were familiarized with caricaturized versions of the same faces, and then asked to recognize their veridical versions, showed no other-race effects on Indian faces. The final experiment sought to determine whether 3D reconstructions from 2D frontal images could be used for the same purpose. Participants who were familiarized with reconstructed faces and then asked to recognize the ground truth versions of the faces showed a significant reduction in performance. In addition, participants who were familiarized with caricatures of reconstructed versions, and then asked to recognize their corresponding ground truth versions, showed a larger reduction in performance. These results are critical for the development of training tools based on computer-generated photorealistic caricatures from "mug shot" images.

DEDICATION

I dedicate this dissertation to my parents for their constant support and encouragement.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

ABSTRACT... iii

DEDICATION v

ACKNOWLEDGEMENTS vi

TABLE OF CONTENTS.. vii

LIST OF FIGURES x

1. INTRODUCTION 1

    1.1 Organization of this document...4

2. BACKGROUND AND LITERATURE REVIEW..6

    2.1 Perceptual effects in face recognition 6
        2.1.1 Facial distinctiveness and the caricature effect..................................6
        2.1.2 Anti-caricatures and anti-faces..10
        2.1.3 The other-race effect 11
        2.1.4 The reverse caricature effect..14
        2.1.5 Reversibility of other-race effects..16
    2.2 Computational models of face perception...18
        2.2.1 Multidimensional face space 18
        2.2.2 Face-space-R...23
        2.2.3 Other-race effect simulations...25
        2.2.4 The other-race effect: the speech analogy 28
        2.2.5 Discussion 31
    2.3 3D face reconstruction 32

3. GENERATION OF 3D CARICATURES 35

    3.1 Three dimensional face models..35
    3.2 Face normalization 38
    3.3 Caricaturization 40

LIST OF FIGURES

1. INTRODUCTION

Law enforcement and security officers must be vigilant in their day-to-day duties of at-large suspects and known terrorists. These officers must study offender's faces to be able to identify and apprehend them. However, learning and recognition of faces are not easy tasks because certain faces are more difficult to recognize than others. Several studies have demonstrated that distinctiveness plays an important role in face recognition by humans. For example, typical faces are more difficult recognize than distinctive faces (Going and Read 1974; Light et al 1979). Furthermore, people have a learned bias toward recognizing faces from their own race better than faces from another race (Valentine and Endo 1992). This other-race effect could negatively affect the accuracy of law enforcement officers recognizing individuals, and may lead to false positive identification.

This research explores ways in which our ability to recognize faces (typical or other race) could be improved through computer-assisted face recognition training tools, specifically through the use of caricaturization. Researchers have demonstrated that caricatures are recognized faster and more accurately than unexaggerated (i.e., veridical) faces (Light et al 1979; Valentine and Bruce 1986b, c). However, is it feasible to use caricatures of unfamiliar individuals as training aid to improve recognition of their veridical faces (i.e., their real and unaltered faces) at a later time? If it is possible,

_____

This dissertation follows the style of *Perception*.

can the same strategy help us improve our recognition of other-race faces? These core questions must be addressed before attempting to implement such face recognition training tool.

The main objective of this dissertation is to answer those core questions. We must determine the extent to which human face recognition of veridical faces improves by using faces whose distinctiveness has been enhanced. To achieve this objective, we developed an automated caricature generation method that uses a database of morphable 3D faces. Then, we performed perceptual studies where participants were familiarized with caricaturized faces and were asked to recognize them in veridical form at a later time. We also studied the extent to which these caricatures may be used to reduce own-race biases.

After establishing the viability of 3D caricatures as a method to improve face recognition, we then addressed the more practical issue of how to obtain 3D facial models from individuals. Returning to the law enforcement scenarios, it is likely that most of the face data comes from "mug-shot" images, where the required 3D information is missing. An alternative in this case would be to reconstruct a 3D model from 2D images prior to the caricaturization process. Though recovering 3D information from 2D images is an ill-posed problem, over the past fifteen years a number of computer methods have been proposed to reconstruct 3D faces from single images, such as the seminal work of Blanz and Vetter on the morphable model (Blanz and Vetter 1999). Although these 3D reconstruction methods produce high quality
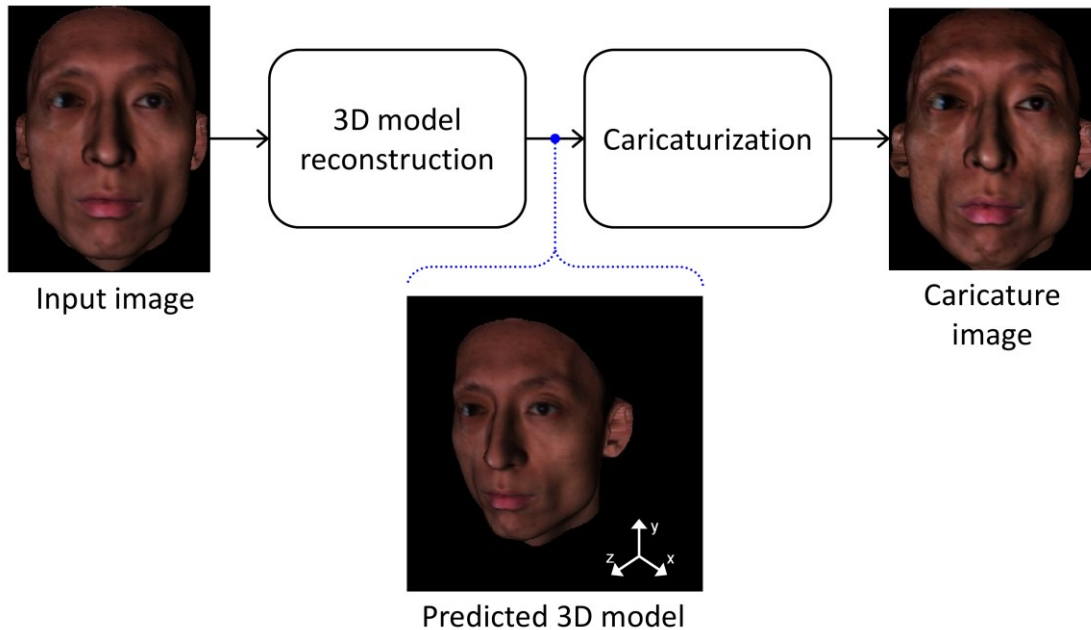
Figure 1. High-level system architecture. The system takes a frontal image as input, and reconstructs a 3D model from the image. The difference between the predicted 3D model and an average 3D face is emphasized, and a caricaturized 3D model is then generated. Finally, a frontal image is rendered based on the 3D caricaturized model.

photorealistic faces, it is unclear whether caricatures from these reconstructed models are effective for training. Thus, we also studied the extent to which caricatures from 3D reconstructed models can help to enhance recognition performance and reduce other-race effects.

Our rationale is that setting up this perceptual foundation will allow us to better define a design framework for a face recognition training tool. This tool could have a broad impact in homeland-security applications, such as training of law enforcement agents to better recognize suspects. Figure 1 shows a high level architecture for the

training tool we envision. This architecture consists of two basic components: 3D model reconstruction, and caricaturization:

- The 3D face reconstruction component processes a 2D frontal image of a target individual, and estimates a 3D facial model using existing reconstruction algorithms, e.g., Blanz and Vetter (1999); Zhang et al (1999); or Moghaddam et al (2003).

- The caricaturization component distorts a 3D model relative to an average 3D face model (Brennan 1985). This 3D average face is calculated based on the system's internal 3D model database. Finally, this component also renders the caricaturized frontal image.

1.1 Organization of this document

The remaining parts of this dissertation are organized as follows. Section 2 provides a review from the perceptual and cognitive psychology literature on the distinctiveness of human faces, caricature effects, and race effects in face recognition by humans. Furthermore, it presents several computational models of face perception. Finally, this section reviews several computer algorithms for 3D face reconstructions. Section 3 presents an overview of the caricature generation process. Section 3 also discusses the 3D face scan database and face normalization procedure that allows us to caricaturize the 3D faces consistently and evenly. Section 4 presents a perceptual study that established proof-of-concept for our working hypothesis, namely that caricaturing can

aid in the memorization and recognition of veridical faces (i.e., the reverse-caricature effect). Section 5 presents a second study that explored the use of reverse-caricature training as a method to reduce other-race effects. These two studies assume that 3D facial models are readily available; Section 6 presents our methodology for generating 3D reconstructed stimuli from 2D images. Section 7 describes the final perceptual study where we determined the extent to which caricatures from 3D face reconstructions are also effective to improve recognition and reduce the own-race bias. Finally, Section 8 presents the conclusions and future directions of this research.

2. BACKGROUND AND LITERATURE REVIEW

This section reviews the most common perceptual effects in human face recognition. It also discusses several computational models based on those perceptual effects. Furthermore, it reviews multiple 3D face reconstruction algorithms from images.

2.1 Perceptual effects in face recognition

It is common to feel that certain faces are harder to remember than others. For example, members from a different racial group than ours seem to look alike, and a caricature of a political figure seems easier to recognize than an actual photograph. These scenarios are well documented in the cognitive psychology literature; the former is known as the other-race effect and the latter is known and the caricature effect. Both and related effect are reviewed in detail in the subsequent sections.

2.1.1 Facial distinctiveness and the caricature effect

Across a wide range of social judgments and perceptions, individuals' distinct qualities typically are given greater weight than their more common or normative qualities (Blanton and Christie 2003). This principle applies to face recognition as well. In general, a prominent chin, nose or hair style can grab attention during initial exposure and then be used as a memory aid during recognition. For instance, a study by Winograd (1981) presented evidence that facial-distinctiveness aids memorization of face photographs. Other studies have shown that distinctive faces are remembered

more accurately (Light et al 1979; Valentine and Bruce 1986b) and faster (Valentine and Bruce 1986c) than typical faces. This simple fact can be harnessed to create more memorable stimuli. Researchers pursuing this strategy create "caricatures" of normal faces by exaggerating their distinct qualities, and they find that people are more able to recognize these distorted faces than the veridical faces that were used to create them. This *caricature effect* is well established in the literature on face recognition (Benson and Perrett 1994; Cheng et al 2000; Lee et al 2000) (Mauro and Kubovy 1992) (Rhodes et al 1987).

Tversky and Baratz (1985) used caricatures of famous individuals created by cartoon artists, and compared them against photographs in memory-recall tasks. The authors reported that photographs were recognized more accurately than caricatures in both recognition[1] and identification[2] tasks, which suggests that distinctiveness does not improve recognition performance. Although these studies are not supportive of a caricature hypothesis, many have argued this is so because photographs contain more spatial information than artistic caricatures (Rhodes et al 1987) or because the artists may have used additional caricaturing techniques (e.g., symbolic caricaturing) in addition to the exaggeration of distinctive features (Mauro and Kubovy 1992). Rhodes et al (1987) suggested that a better comparison of distinctiveness effects should be made

_____

[1] Recognition tasks are usually old/new test.
[2] Identification tasks are usually name/object association test.

8

8

| (a) | (b) | (c) |
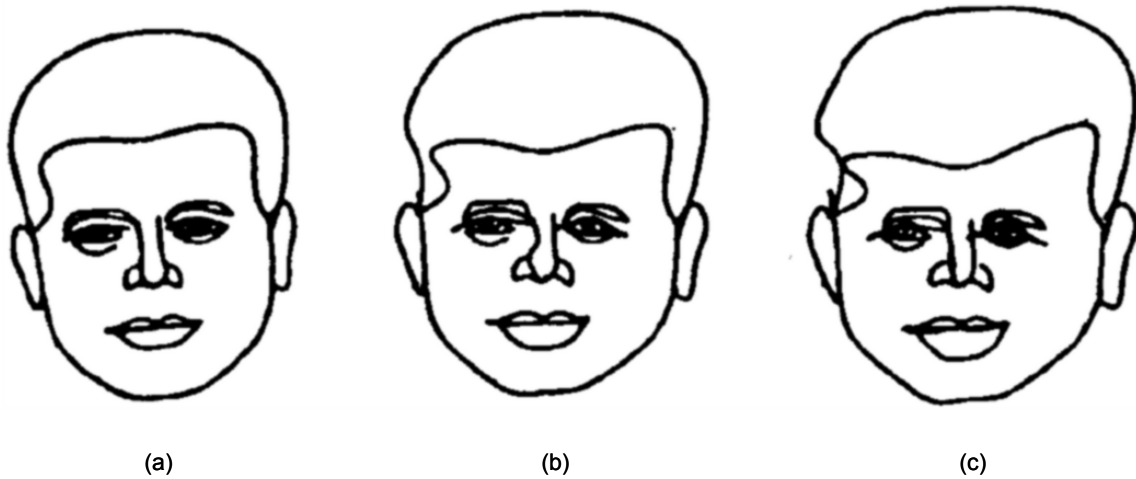
Figure 2. An example produced by the Brennan's caricature generator. John F. Kennedy: (a) $\alpha =$ 0% exaggeration, (b) $\alpha = 50\%$ exaggeration, (c) $\alpha = 100\%$ exaggeration. This figure was obtained from Rhodes et al (1987).

within the same representational medium.

Rhodes et al (1987) used a computer-based caricature generation tool (Brennan 1985) to create line-drawings of faces based on 169 facial points. An average face (f_{avg}), was computed, and a caricature (f_C) of a given face (f) were generated by exaggerating the difference between the points of a given face and their correspondences on the average face by a factor α (see Equation (1) and Figure 2). Identification tests of familiar

$$f_C = f_{avg} + (1 + \alpha) \cdot (f - f_{avg}) \tag{1}$$

faces using veridical and caricaturized line-drawings indicated that caricatures were identified faster than veridical line-drawings but not more accurately. This result led the authors to propose two possible explanations for the caricature advantage.

According to the first view, faces would be represented in memory as caricatures instead of veridical faces. The second view hypothesizes that memory representations are veridical but caricatures facilitate the access of the veridical faces. Benson and Perrett (1994) performed similar experiments using an improved line-drawing caricature generation tool (Benson and Perrett 1991b); they also reported that caricatures were identified *faster*, but also found a caricature advantage over veridical line-drawings in terms of recognition *accuracy*. Stevenage (1995) and Mauro and Kubovy (1992) successfully demonstrated the caricature advantage over veridical representations using hand drawings and an Indentikit[3] system, respectively. Stevenage (1995) also compared caricature line-drawings against veridical line-drawings during training to assess whether they aid in recognizing photographic faces. Her study showed a caricature advantage over the veridical line-drawing when the amount of training was unlimited, but caricature performance was as good as the veridical when the amount of training was restricted.

The caricature advantage has not been restricted to line drawings only. Benson and Perrett (1991a) have also shown that photographic-quality caricatures improve recognition of natural facial images as well. These authors borrowed Brennan's technique (Brennan 1985), and extended it to generate caricatures on pixel-based

_____

[3] Identikits are composite sketches of faces based on verbal descriptions commonly used by law enforcement agencies.
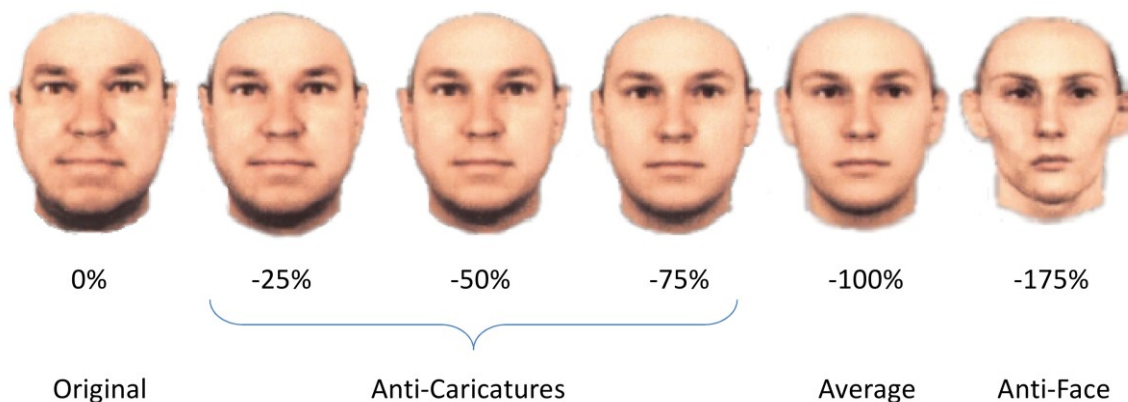
Figure 3. Examples showing an original face (α = %0), anti-caricatures (α = -25%, -50%, -75%), the average face (α = -100%), and the corresponding anti-face (α = -175%). Faces were obtained from (Leopold et al 2001).

images. Their results indicated that the caricature advantage in photos does indeed exist, but at a lower caricaturization level than the one used by Rhodes et al (1987) for line-drawing caricatures: the authors reported a disadvantage when using the caricaturization levels of Rhodes et al.

2.1.2 Anti-caricatures and anti-faces

Rhodes et al (1987) also studied face recognition performance on anti-caricatures. Anti-caricatures can be created by reducing differences between an original face and a norm face, see Figure 3 for α = -25% thru -75% (Benson and Perrett 1991a; Rhodes et al 1987; Rhodes and McLean 1990; Rhodes and Tremewan 1994; Stevenage 1995). Rhodes et al (1987) found that anti-caricatures of line drawings led to significantly slower reaction times than veridical and caricaturized faces. However, Rhodes et al (1987) did not find a significant difference in recognition accuracy between veridical and anti-caricaturized

faces. Rhodes and McLean (1990) found that anti-caricatures had significant slower reaction time and worst recognition accuracy than veridical stimuli. However, Rhodes and McLean (1990) did not use faces for their studies, they use line drawing of birds as their homogeneous stimuli. Benson (1991a) obtained similar results using photographic faces, namely, that anti-caricatures were recognized worse than veridical or caricaturized faces.

Anti-faces have opposite features to an original face, see Figure 3 for α = -175% (Jiang et al 2006; Leopold et al 2001). Namely, anti-faces lie on the other side of the mean in face space (Blanz et al 2000). People identify a veridical face more accurately if they see it immediately after viewing its corresponding anti-face (Leopold et al 2001).

2.1.3 The other-race effect

Face recognition has also been studied in the context of race effects (Chance and Goldstein 1981; Goldstein and Chance 1980; Lindsay et al 1991; Valentine and Bruce 1986a; Valentine and Endo 1992). These studies show that subjects tend to recognize faces of their own race better than faces from another race (Brigham and Malpass 1985; Chance and Goldstein 1996; Shepherd 1981). This observation is also known as the *own-race bias* or *other-race effect*, which largely reflects race-related differences in perceptual expertise (Brigham and Malpass 1985; Lindsay et al 1991) although other factors could be implicated in this effect, like physiognomic homogeneity, or ethnic attitudes (Chance and Goldstein 1996; Malpass 1981). Researchers have also noted the

robustness of this effect, which has been reproduced across many studies (see Bothwell et al 1989; Meissner and Brigham 2001; Shapiro and Penrod 1986) and across racial groups (Bothwell et al 1989; Brigham et al 2007; Kassin et al 1989). We will survey some these studies and discuss the potential factors causing the other-race effect.

To explain other-race effects, one may be tempted to infer that faces from different races are inherently more difficult to recognize that others. A number of studies, however, have concluded that there is no significant difference in variability within faces of different races (Goldstein 1979a, b; Goldstein and Chance 1976; Shepherd and Deregowski 1981). It has also been suggested that interracial attitudes, in which subjects may have more favorable attitude towards members of their own race, may be a factor in other-race effects (Malpass 1981). However, no evidence has been found to associate interracial attitudes and face recognition performance (Brigham and Barkowitz 1978; Lavrakas et al 1976).

A more likely factor which contributes to the production of the other-race effect is level contact (or lack thereof) with members of the other race. Individuals have more experience with members of their own race; therefore they are better equipped at recognizing them (Chance et al 1975; Goodman et al 2007; Malpass 1981; Vrij and Winkel 1989). A number of studies supporting this contact hypothesis can be found in the literature. Lindsay et al (1991) studied the other-race effect by performing a perceptual task using Caucasian and African American subjects. The authors found that Caucasian subjects performed worse in recognizing African American faces than in

recognizing Caucasian faces, but African American subjects performed equally well in recognizing faces from both groups. Subjects were also surveyed about their interaction with the other race to determine a level of expertise; the results showed that African American subjects had higher level of interaction (i.e. contact) with the other race than Caucasian subjects. The authors reported a positive correlation between recognition performance for other-race faces and the amount of interaction with members of the other race. Another study (Valentine and Endo 1992) evaluated the effects of distinctiveness between different races, where British and Japanese subjects were asked to rate the distinctiveness of faces from their own race and from the other race. The correlation between the ratings assigned by British subjects, and the ratings assigned by Japanese subjects on British faces was 0.82, while the correlation on Japanese faces was 0.65. According to the authors, these results indicate that the perception of distinctiveness varies depending of the level of familiarity with faces of a particular race. The authors also performed old/new face recognition experiments using distinctive and typical faces from Japanese and British individuals. The experiments showed that British subjects recognized British faces significantly better than Japanese faces, but Japanese did not show own-race bias effects. The authors indicated the possibility that Japanese subjects were more familiar recognizing Caucasian faces than Caucasian subjects with Japanese faces, due to the recent influence of Western films and TV (Valentine and Endo 1992).

Other-race effects may be rooted to the fact that the most appropriate features for discriminating own-race faces are not diagnostic for other-race faces (Lindsay et al 1991; Shepherd and Deregowski 1981). For instance, Ellis et al (1975) used verbal descriptions to study the frequency of facial features that were used describing Caucasian and African faces by both Caucasian and African participants. The authors showed that Africans focus more on the shape and location of the eyes, eyebrows and ears, whereas Caucasians focus more on hair texture, and hair and eye color (Ellis et al 1975).

As presented above, familiarization plays a role in our ability to recognize faces, specifically faces from another race. These previous studies have shown that distinctive faces are better recognized than typical faces, and that our perception of distinctiveness is a function of our level of familiarity with a given group of faces. Collectively, these results support the use of caricatures as training aid to familiarize individuals with new faces, specifically faces from another race.

2.1.4 The reverse caricature effect

A handful of studies have demonstrated a *reverse-caricature effect* in which familiarization with caricatures that highlight the distinct features of a face (i.e., those aspects that are uncommon) improves recognition of an image of the veridical face at a later time (e.g., Benson and Perrett 1994; Cheng et al 2000; Deffenbacher et al 2000; Lee et al 2000; Mauro and Kubovy 1992; Rhodes et al 1987; Stevenage 1995). This reverse-

caricature effect highlights the potential for the use of caricatures for training in applied settings[4].

A variety of issues surrounding the reverse-caricature effect need to be clarified before the technique will be ready for use in an applied setting. For example, due to technological limitations, the earliest studies of the caricature effect (both direct and reverse) involved some combination of caricaturized photographs or line drawings (Benson and Perrett 1991a; Lee et al 2000; Mauro and Kubovy 1992; Rhodes et al 1987; Stevenage 1995). In more recent work (e.g., Deffenbacher et al 2000; O'Toole et al 1997) researchers have used three-dimensional (3-D) laser scans that, together with the application of a caricature algorithm (see Benson and Perrett 1991b; Brennan 1985; Rhodes et al 1987), allow for much greater experimental control than can be obtained using photograph manipulations or line drawings. However, the influence of this new technology, as well as and other presentation choices one might make, have not been studied in a systematic fashion.

In one demonstration of the reverse-caricature effect, Deffenbacher et al (2000) used presentation of a three-quarter view of the 3-D stimuli, at both familiarization and

_____

[4] A conceptual parallel to the reverse-caricature effect can be found in speech recognition, where exaggeration of auditory features can aid in learning unexaggerated nonnative phonetic contrasts (McCandliss et al 2002). Consider as an example the difficulty that native speakers of Japanese have trying to discern the difference between the English phonemes [r] and [l] (this occurs because this particular contrast in sounds is not present in Japanese.) Research indicates that training with spectrally-exaggerated versions of these two sounds can help Japanese speakers recognize the unexaggerated versions of these same sounds (McCandliss et al 2002).

test, and this choice of presentation may have been consequential. Other studies focusing on veridical faces have demonstrated that there is an advantage in recognition memory for faces that employ a three-quarter-view representation, when compared to either a frontal or profile representation (e.g., Bruce et al 1987; Krouse 1981; Logie et al 1987). It remains unclear, however, whether the reverse-caricature effect will generalize to 3-D scans in which caricatures are presented only frontally during familiarization. Also of note, Deffenbacher et al (2000) familiarized participants with target faces for a total of 60 seconds each prior to recognition testing. The authors indicate that a reverse caricature effect does not occur with shorter familiarization periods, a constraint that could undermine the feasibility of reverse-caricaturing as a training approach in practical scenarios.

2.1.5 Reversibility of other-race effects

The other-race effect has been studied extensively on the past three decades. Furthermore, there are studies which show that our face recognition system has the plasticity to reverse the other-race effect (Goodman et al 2007; Sangrigoli et al 2005), making training methods to reduce this effect feasible. However, there are only few studies on face recognition training, specifically attempting to remove the other-race effect (Malpass 1981).

Sangrigoli et al (2005) tested East Asian adults adopted by European Caucasian families when they were between 3 and 9 years old, versus East Asian natives. The

authors showed that adoptees did not have the other-race effect on Caucasian faces, while the East Asian natives showed a significant other-race effect. Notice that the other-race effect appears to develop between the ages of three and five (Pezdek et al 2003).

Lavrakas et al (1976) showed immediate reduction of the other-race effect after training Caucasian subjects with African faces using concept learning[5] tasks. In this study, Caucasian participants were asked to identify the concept feature (e.g., light eyes) in a series of African faces. Immediately after the learning stage, participants showed a reduction of the other-race effect in a recognition task. However, a recognition task performed a week after the learning stage did not show any reduction of other-race effects. In contrast, however, an earlier study by Malpass et al (1973) failed to show a reduction of the other-race effect when using verbal training tasks.

Hills and Lewis (2006) showed that other-race effects could be reduced by familiarizing subjects with own-race faces containing features critical for differentiating other-race faces. This suggests that the other-race effect may be reduced by drawing attention to the most distinctive feature of a given face, potentially overriding the own-

_____

[5] Concept learning is a strategy which requires a learner to compare and contrast groups or categories that contain concept-relevant features with groups or categories that do not contain concept-relevant features. Such concepts are mental representations which enable one to discriminate between objects that satisfy the concept and those which do not (Goodman et al 2008).

race feature biases of the viewer. Hills and Lewis' results led us to pose these questions:

(*i*)   Can caricatures be used as a more general approach to direct attention to critical features for the identification of other-race faces?

(*ii*)   Can we use the reverse-caricature effect to reduce the other-race effect?

2.2 Computational models of face perception

This section introduces the multidimensional face space framework of Valentine (1991), an important model that has been vastly used to formulate how we store and perceived faces. This space encodes features that best identify a particular face and accounts for multiple perceptual effects, such as distinctiveness, caricature, and other-race effects. Other models discussed here have been based on Valentine's face space.

2.2.1 Multidimensional face space

The hypothesis that memory representations are based on facial distinctiveness assumes that there is a norm or prototype face that is typical in a given population (Rhodes et al 1987; Valentine 1990, 1991). A norm-based model assesses the distinctiveness of a face using a similarity measure relative to a norm or prototype face. Valentine (1991) developed a theoretical framework which is consistent with both the norm-based model and the experimental results obtained to that date. Valentine also included in his framework an exemplar-based model, which assessed the distinctiveness independently from a norm face. An understanding of these models
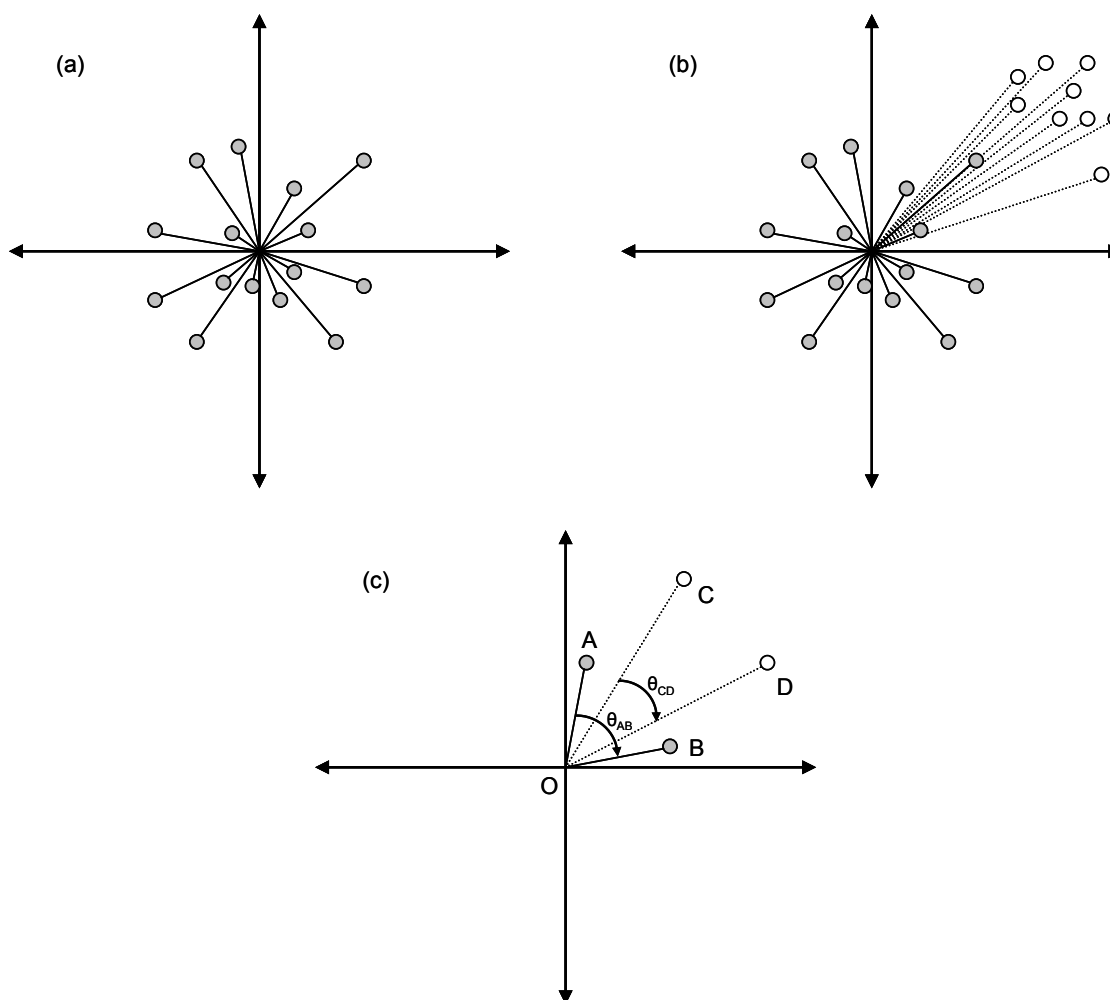
Figure 4. Norm-based coding. (a) shows a two-dimensional representation of faces. The origin represents norm or prototype face and faces are defined as vectors; (b) includes the other-race faces; (c) illustrates the effect of similarity on the norm-based coding. These figures were adapted from Valentine (1990).

may help develop better caricature generation methods, create simulations that account

for the current perceptual results, or improve automatic face recognition algorithms.

According to the multidimensional face space framework (MDFS) (Valentine

1990, 1991; Valentine and Endo 1992), which assumes a norm-based model, faces are

encoded as vectors in a multidimensional space in which the norm or prototype face is

located at the origin (Figure 4a). Faces in this space are assumed to be normally distributed around the norm face, so regions closer to the origin are more densely populated than those away from the norm. Under the norm-based model, the similarity between two faces is defined as a function of the inner-product of their two vectors. During a recognition task, this similarity is computed between the stimulus face (i.e., test face) and the faces encoded in this multidimensional space. The stimulus face is recognized if the closest encoded face is the target face. If one generates a stimulus face that is close to the target face (e.g., the target face under slightly different illumination conditions), and the target face is located in the low density region of the multidimensional space, the probability of the target being selected as a match is higher than if the target was in the high density region. This explanation is consistent with the result that distinctive faces are recognized more accurately than typical faces.

The above model assumes that all faces are from the same race. Can the norm-based model account for race effects between a given race (own-race) and another race (other-race)? If one includes faces from another race while maintaining the own-race norm at the origin, the other-race faces form a cluster with their own prototype and distribution away from the origin (Figure 4b). The other-race face vectors will have higher similarities among themselves than the own-race face vectors, making the other-race faces harder to discriminate than the own-faces. Figure 4c illustrates this similarity difference between own-race (points A and B) other-race faces (points C and D) when using the inner-product metric. In this figure the similarity between points C and D is

higher than the similarity between points A and B. But since the norm-based coding assumes that the norm to encode the faces is the own-race norm, there is a conceptual problem when one tries to account for distinctiveness within the other-race faces in this model (Valentine and Endo 1992). Assuming that the other-race faces are normally distributed around their central tendency, distinctive other-race faces will be further away from this other-race central tendency and typical other-race faces will be closer. Thus, the vector representation relative to the own-race norm (Figure 4b) is inappropriate to capture this normal distribution of distinctiveness within other-race faces because the model's similarity metric is based on the inner product. These other-race vectors could not be used to properly compute the similarity within the other-race faces as one would within the own-race faces. In consequence, in order to account for distinctiveness with other-race faces, a model should not rely on an own-race norm when computing a measure of similarity between faces.

Based on this, Valentine and Endo (1992) suggested an exemplar-based coding as a better model for the perceptual results on distinctiveness and race effects. The exemplar-based model assumes that each face example is stored independent of a norm face as a point in the multidimensional face space (Figure 5a). Like the norm-based coding, this model assumes that the encoded faces are normally distributed. Therefore, typical faces are in an area of high exemplar density, whereas distinctive faces are in areas of low exemplar density. Distinctiveness effects are explained using exemplar density, as in the norm-based coding. The only difference between both models is in
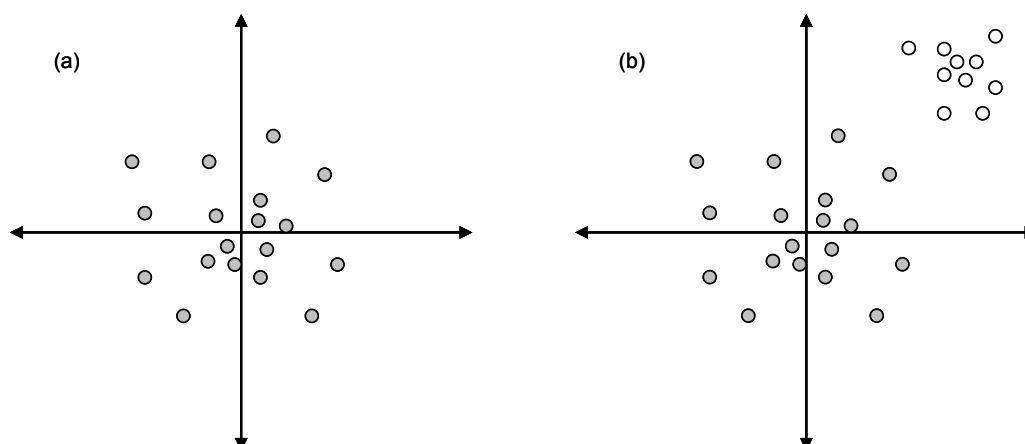
Figure 5. Exemplar-based coding. (a) shows a two dimensional representation of faces. The axes origin was arbitrarily placed at the central tendency of the population; (b) shows the other race faces. Notice that the other-race are more densely clustered. These figures were adapted from Valentine and Endo (1992).

the similarity measure: the exemplar-based coding uses the Euclidean distance between points. Hence, a distinctive face is recognized more accurately because it will be farther from its nearest neighbor than a typical face would be from its nearest neighbor.

Valentine's exemplar-based model accounts for the other-race effect by assuming that other-race face points are more densely clustered than the own-race faces (Figure 5b). The evidence to support this assumption comes from the fact that the dimensions more appropriate to differentiate own-race faces might not be the same ones for the other-race faces (Shepherd and Deregowski 1981), e.g., Africans may focus more on the shape and location of the eyes, eyebrows and ears, whereas Caucasians may focus more on hair texture, and hair and eye color (Ellis et al 1975; Lindsay et al 1991). Using this density difference between own-race and the other-race faces, the own-race face will be recognized more accurately than a face from another race. Since

the Euclidean distance measure also accounts for the distinctiveness of the other-race, the exemplar-based model does not suffer the problems encountered by norm-based coding.

2.2.2 Face-space-R

Lewis (2004) proposed a unified model for face recognition that accounts for distinctiveness, caricature and race effects. His model is an updated version Valentine's exemplar-based model, but accounts for the effects of proximity of all other exemplars (Lewis 2004). In contrast to previous face space representations, Lewis's model is explicitly defined. The model is based on an activation response function $r_i(x)$ defined as follows:

$$r_i(x) = act_i(x) \cdot \lambda_i - \sum_{\forall j \neq i} \{act_j(x) \cdot \lambda_j\} \tag{2}$$

where $act_i(x)$ is the activation of a probe face x to a given gallery face f_i and defined:

$$act_i(x) = e^{\frac{-\|x - f_i\|^2}{2\sigma^2}}. \tag{3}$$

The parameter λ_i incorporates level familiarity of a face i into the model. At the boundary, $\lambda_i = 0$ is an exemplar never encountered before, and $\lambda_i = 1$ represens a highly familiar face. Finally, σ defines the error, where smaller σ represents a good

quality view of a face and larger σ represents a blurred, brief or bad angle views of a face. The second term in $r_i(x)$ accounts for total activation of the remaining faces in the face space. The activation response $r_i(x)$ of a particular exemplar is determined by that exemplar's activation minus the activation of all other exemplars. So the exemplar with a positive response is the match identity of the probe. If none of the exemplars produces a positive response then the probe is considered an unfamiliar face. The free parameters of the computational model are, (*i*) the number of dimensions, (*ii*) the spread σ or error, and (*iii*) the number of faces the model can recognized. These free parameters are setup such that a caricature probe would have higher activation response than the corresponding veridical probe.

The model obtains higher activation for faces away from the mean, as the face space is sparser, which is consistent with the distinctiveness effect. The model is also consistent with the effect that distinctive faces require less caricaturization than typical faces to achieve the same level of likeness. Finally, the model accounts for the other-race effect. If the model includes other-race exemplars away from the central tendency of the face space (own-race mean), the average activation response is higher for own-race probes than the average one for the other-race probes.

Although this model can account for distinctiveness, caricature, and race effects, it does not account for reverse caricature effects. This is due to the nature of the models free parameter estimation, which assumes that the activation response for caricature is higher than veridical.

2.2.3 Other-race effect simulations

A few researchers have built computational models that attempt to reproduce the perceptual results of caricature and race effects (Caldara and Abdi 2006; Furl et al 2002; Haque and Cottrell 2005; Lewis 2004; O'Toole et al 1991; O'Toole et al 1994; Tanaka and Simon 1996; Valentin et al 1997). Furl et al (2002) evaluated a number of face recognition algorithms to study the extent to which they exhibited OREs. The authors found that algorithms in which the face space captures the lifetime knowledge of a given category of faces (e.g., a race, age group), show significant other-race effect. Caldara and Abdi (2006) also simulated the ORE using a biased auto-associative network. Results from this study showed a significantly larger classification error for gender in the minority race than in the majority race. O'Toole et al (1991) simulated the other-race effects using auto-associate networks (Hopfield 1982). The "own-race bias" was achieved by training the network with a majority of faces from one race and a small number of faces from another race. This bias led the network to capture those features that were more appropriate to the majority faces, making the network results consistent with the exemplar-based coding model. Following training, the network was used to generate majority new faces and minority new faces by presenting novel images from both groups and recording the recalled faces. Using a similarity analysis, it was found that the generated majority faces were more distinctive than the generated minority faces. This result is consistent with the other-race perceptual effect.

In more depth, O'Toole et al (1991) followed a two-stage memory model that loosely resembles human memory models (see Atkinson and Shiffrin 1968). This memory model contains a long-term component that represents the lifetime knowledge of faces (i.e. primarily own-race faces) and a short-term component representing recently learned faces (i.e. a mix of own-race and other-race faces). O'Toole et al (1991) trained both components using an auto-associative network with error-correction – essentially an on-line implementation of principal components analysis (PCA). In this study, the long-term component was biased toward a single (majority) race, whereas the short-term component had an equal amount of faces from two different races (e.g. majority and minority). Using PCA, a long-term face $f_L \in L$ and a short-term face $f_S \in S$ was described as follows:

$$f_L = l_{avg} + \sum_{i=1}^{m-1} \alpha_i \cdot l_i \tag{4}$$

$$f_S = s_{avg} + \sum_{i=1}^{n-1} \beta_i \cdot s_i \tag{5}$$

where l_i and s_i denote the long-term and short-term eigenvectors, l_{avg} and s_{avg} are the average long-term and short-term face, α_i and β_i are the corresponding principal components of a particular long-term and short-term face, and m, n are the number of long-term and short-term faces, respectively.

O'Toole et al proposed a recognition task that is analogous to the standard old/new recognition paradigm (see Goldstein and Chance 1980), where human participants are asked to memorize a set of faces (the OLD faces), and later are asked to recognized them among a set of confounders (the NEW faces). Here, the short-term faces S are treated as the OLD faces and a separate set of faces $N \notin \{L \cup S\}$ are considered as the NEW faces. Recognition is achieved by computing the cosine of the angle between each test face $f_{test} \in \{S \cup N\}$ and its reconstruction f'_{test}, and comparing it against a decision threshold E:

$$\varepsilon = \frac{f'_{test} \cdot f_{test}}{\|f'_{test}\|\|f_{test}\|} \begin{cases} > E & f_{test} \leftarrow OLD \\ \leq E & f_{test} \leftarrow NEW \end{cases} \tag{6}$$

and the reconstructed test face f'_{test} is obtained by projecting the original face f_{test} into each memory space independently, and then weighting each resulting reconstruction. Namely, the principal components of f_{test} on the long- and short-term face space are computed as follow:

$$\alpha'_i = \left(f_{test} - l_{avg}\right)^\top \cdot l_i; \ \forall i \in \{1, 2, \dots, m-1\} \tag{7}$$

$$\beta'_i = \left(f_{test} - s_{avg}\right)^\top \cdot s_i; \ \forall i \in \{1, 2, \dots, n-1\} \tag{8}$$

from which the two reconstructions $f'_{test(L)}$ and $f'_{test(S)}$ are obtained as:

$$f'_{test(L)} = l_{avg} + \sum_{i=1}^{m-1} \alpha'_i \cdot l_i \tag{9}$$

$$f'_{test(S)} = s_{avg} + \sum_{i=1}^{n-1} \beta'_i \cdot s_i \tag{10}$$

which are finally weighted to obtain the final overall reconstruction f'_{test}:

$$f'_{test} = w_l \cdot f'_{test(L)} + (1 - w_l) \cdot f'_{test(S)} \tag{11}$$

where w_l is a long-term weighting factor, set to a value of 0.75 in the original study by O'Toole et al (1991). Even though the study successfully reproduced OREs, the model in equation (11) raises a few questions. First, it is unrealistic to assume that a lifetime of experience with own-race faces can be altered by recently learned faces –at least not by as much as 25%, as the study suggests. Second, because the short-term memory contains an equal number of own-race and other-race faces, the net effect is a rebalancing of both classes, which by necessity will reduce OREs.

2.2.4 The other-race effect: the speech analogy

In the context of speech recognition, McClelland (2001) used Kohonen's Self Organizing Maps (SOM) (Kohonen 1990) and Hebbian learning[6] to simulate the perceptual problem

---

[6] Hebb's proposed mechanism of learning states that if a neuron excites another neuron repeatedly the connection between them is strengthened.
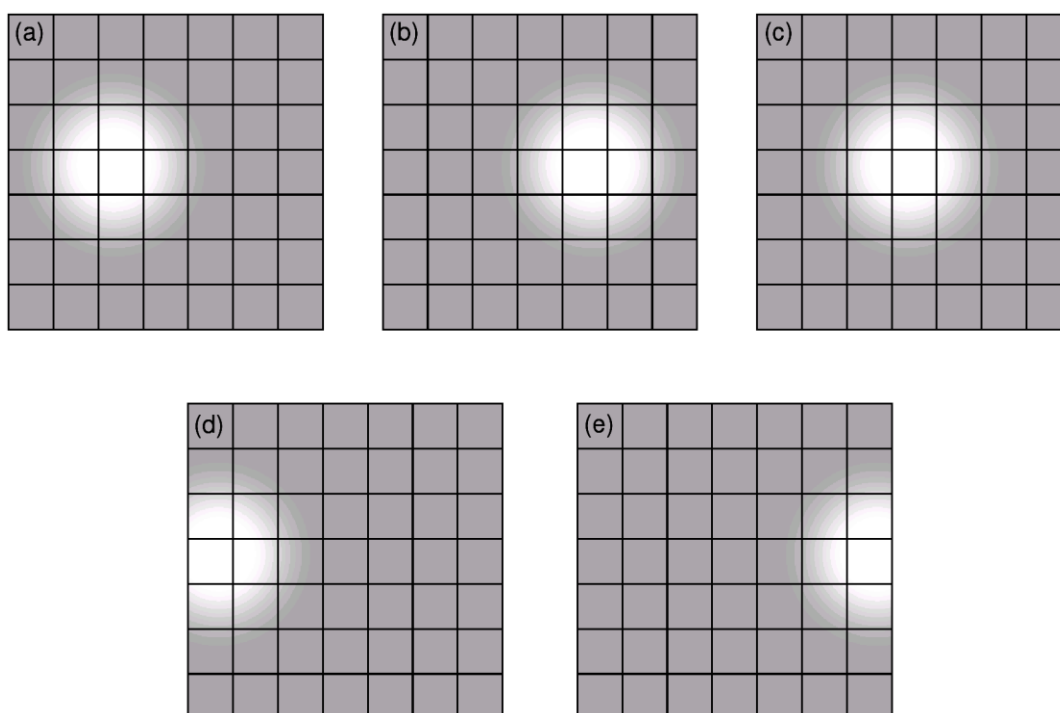
Figure 6. Prototype input patterns for a 7x7 SOM. (a) and (b) show the hypothetical perceptual patterns for English /r/ and /l/ phonemes. (c) shows the corresponding single central prototype for Japanese. (d) and (e) illustrate English /r/ and /l/ patterns exaggerated away from each other. This figure was adapted from McClelland (2001).

that Japanese native speakers encounter when differentiating /r/ and /l/ phonemes in English. For Japanese speakers, those two phonemes are perceived as having the same sound. In consistency with the Hebbian learning rule, McClelland suggested that listening either /r/ or /l/ phonemes would reinforce the Japanese single representation of the phonemes.

For the simulation McClelland represented the /r/ and /l/ phonemes in English as two overlapping hypothetical patterns (Figure 6a, b), whereas in Japanese the phonemes were represented as one central pattern (Figure 6c). After training a group of
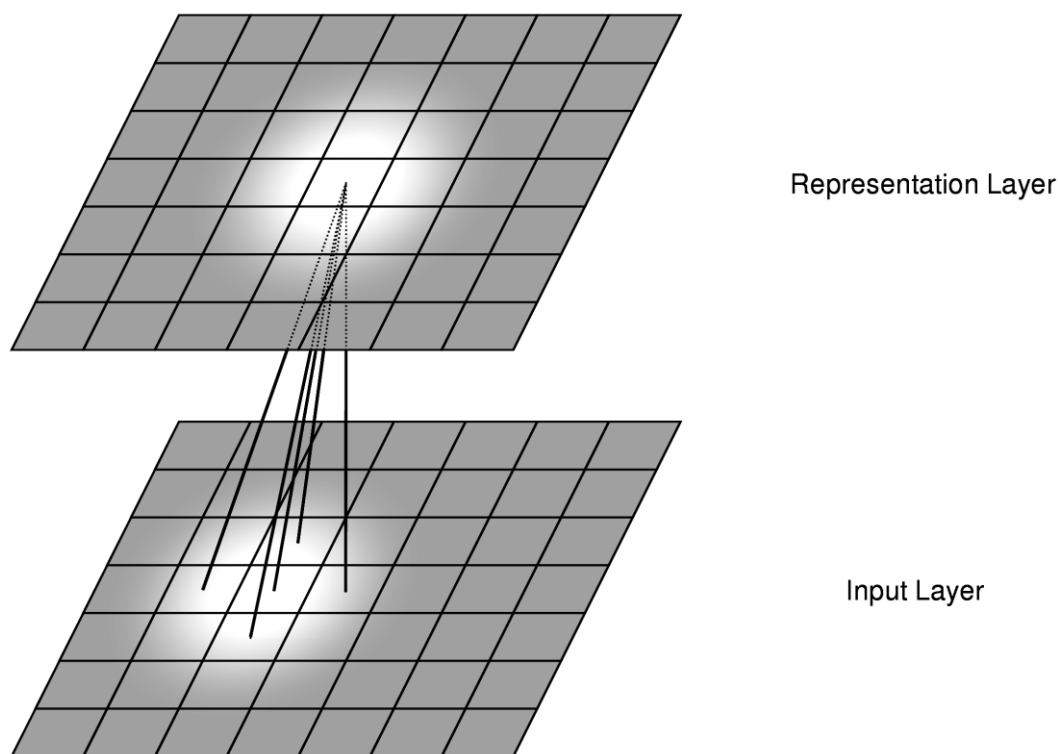
Figure 7. Illustrates the output representation of a "Japanese" trained SOM after presenting one of the "English" overlapping patterns at the input layer. This figure was adapted from McClelland (2001).

SOMs with the Japanese pattern, the English patterns were fed to the networks. Even though the patterns were presented for a large number of epochs, the SOMs fused the English patterns into the Japanese pattern at the output (Figure 7). McClelland pondered whether exaggerating the difference the /r/ and /l/ patterns in the input would activate different output representations. McClelland proceeded by training with exaggerated versions of the English patterns (Figure 6d, e) until the SOMs were able to represent the two patterns as separate. As a result, when the original English patterns were input at a later time, the SOMs were able to separate them as well.

Although, this simulation was quite simple, it shows the benefits that distinctiveness enhancements may provide to pattern classification algorithms.

2.2.5 Discussion

The above studies provide important theoretical insights about several face recognition perceptual effects, such as distinctiveness, caricature, and other-race effects. Furthermore, they allow us to formulate theories about how we represent, store and retrieve facial information. Valentine's framework is by far the most influential model. It accounts for all the effects mentioned above, and also suggests an exemplar-based face space representation. However, these models do not address questions that could have greater practical implications for the design of training tools for human viewers and better automatic face recognition algorithms. How can we use the caricature effect to improve face recognition? To what extent can other-race effects be reversed? What strategies may be used to effectively minimize OREs?

As shown above, simulation of some of the perceptual results on face recognition, such as the race effects or the benefits of enhancing distinctiveness, is possible. However, this area of study has not been explored very extensively, particularly whether caricaturization can aid in the context of automatic face recognition.

2.3   3D face reconstruction

Three dimensional (3D) face reconstructions from images has been an active research problem given its relevance to many fields. 3D face reconstructions have applications in animation, video conferencing, multimedia (Fanany and Kumazawa 2002; Mao et al 2008; Marques and Costeira 2008; Suen et al 2007; Yip and Jin 2005), biometric and security (e.g., face recognition and authentication) (Amin and Gillies 2007; Widanagamaachchi and Dharmaratne 2008), morphing, and games (Widanagamaachchi and Dharmaratne 2008). These reconstructions are an alternative to more expensive methods such 3D scanning (Mao et al 2008; Suen et al 2007). Furthermore, 3D face reconstructions could solve common problems on image-based automatic face recognition systems, as these systems are sensitive to variations in pose, illumination and expressions (Guan 2007; Hu et al 2004; Mao et al 2008; Moghaddam et al 2003). Additionally, images could have missing spatial information due to noise, occlusions, and shadows (Hwang et al 2000). The advantage of having a facial 3D model is that the recognition system can test several settings and conditions such as changes in pose and illuminations (Amin and Gillies 2007; Mao et al 2008; Moghaddam et al 2003).

3D face reconstruction from images is a challenging problem due to the complexity of the face shape, and the subtle and spatially varying reflectance property of the skin (Fanany et al 2002). Reconstructions from a single image are even more problematic, since different shapes may give rise to the same intensity pattern (Hassner

and Basri 2006); however, there is a tendency among the latest face reconstructions methods to use a single image (Fan et al 2010).

There are several methods for 3D face reconstruction, which can be categorized into four major groups: (*i*) shape from shading, (*ii*) shape from silhouettes, (*iii*) shape from motion, and (*iv*) analysis by synthesis (Amin and Gillies 2007; Ming and Ruan 2010). The *shape from shading* approach recovers 3D depth information from a gradual variation of shading in the image (Zhang et al 1999). In contrast, *shape from silhouettes* deals with obtaining the 3D information from outlines of the object in the target images (Cheng and Lai 2001; Matusik et al 2000; Moghaddam et al 2003; Szeliski 1993); in this case, the accuracy of the 3D model depends on the number of images with different views of the object used to perform the reconstruction. *Shape from motion* deals with recovery of 3D points on an object from 2D correspondences of points across images (Amin and Gillies 2007; Cheng and Lai 2001; Choi and Hwang 2002). Finally, *analysis by synthesis* uses a fully aligned 3D face model database to estimate the statistical parameters (e.g., principal components) that better match the projected face shape with the target face image (Amin and Gillies 2007; Blanz and Vetter 1999). There are also methods that use a combination of 3D reconstruction categories mentioned above. For example, Cheng and Lai (2001) use shape from motion and shape from silhouettes to reconstruct 3D faces from a video sequence, Moghaddam et al (2003) use shape from silhouettes and analysis by synthesis to recover the 3D shape of a face, and Atick et al

(1997) use a shape from shading approach in combination with an analysis by synthesis method for their face reconstruction algorithm.

Among all approaches for 3D face reconstruction from images, analysis by synthesis is the most successful method today (Amin and Gillies 2007). Analysis by synthesis is the focus of our attention on later major sections.

3. GENERATION OF 3D CARICATURES

This section describes the caricature generation process that was used in our work. We discuss how 3D face models are defined in the 3DFS-100 database (Blanz and Vetter 1999), present the normalization process, which accounts for the original facial distinctiveness prior to caricaturization, and describe the caricaturization process, which is based on a data-driven caricaturization factor.

3.1 Three dimensional face models

The 3D face models used in this dissertation come from the 3DFS-100 database. This dataset is also used by 3D morphable model fitting algorithm (Blanz and Vetter 1999; Blanz and Vetter 2003). In this analysis-by-synthesis approach, faces (with full correspondence) are stored in terms of two separate shape S_i and texture T_i vectors:

$$S_i = (x_{i1}, y_{i1}, z_{i1}, x_{i2}, \ldots, x_{in}, y_{in}, z_{in})^T \tag{12}$$

$$T_i = (R_{i1}, G_{i1}, B_{i1}, R_{i2}, \ldots, R_{in}, G_{in}, B_{in})^T \tag{13}$$

where i is the index of a particular face and n is the number of vertices in the 3D model. For a database of $i = 1$ to m faces, Blanz and Vetter (1999) compute the average of S and the average of T and performed Principal Component Analysis (see Duda et al 2001; Fukunaga 1990) on the shape matrix $A = (a_1, a_2, \ldots, a_m)$, where $a_i = S_i - s_{avg}$ is a mean-centered shape vector and the texture matrix $B = (b_1, b_2, \ldots, b_m)$, where $b_i = T_i - t_{avg}$ is

a mean-centered texture vector. From this, the shape and texture of a given face are defined by:

$$S = s_{avg} + \sum_{i=1}^{m-1} \alpha_i \cdot s_i \tag{14}$$

$$T = t_{avg} + \sum_{i=1}^{m-1} \beta_i \cdot t_i \tag{15}$$

where α_i are the shape principal components, s_i are shape eigenvectors, β_i are the texture principal components, and t_i are the texture eigenvectors.

To estimate a 3D model, Blanz and Vetter (1999) find the shape and texture principal components in (14) and (15) that minimize the sum-squared error E

$$E = \sum_{x,y} \left\| I_{input}(x,y) - I_{model}(x,y) \right\|^2 \tag{16}$$

where I_{input} is the 2D input image and I_{model} is the 2D rendering of the 3D model. In other words, their algorithm finds the principal components coefficients such that the 2D image rendered from the 3D model is as close as possible to the 2D input image (see Figure 8). Their optimization algorithm uses a stochastic version of Newton's method which only uses 40 randomly selected triangles from the 3D face model to boost speed and to avoid local minima. This fitting algorithm analytically computes the gradient of

$$\alpha_1 \times \qquad + \alpha_2 \times \qquad + \alpha_3 \times \qquad + \ldots$$

$$\beta_1 \times \qquad + \beta_2 \times \qquad + \beta_3 \times \qquad + \ldots$$
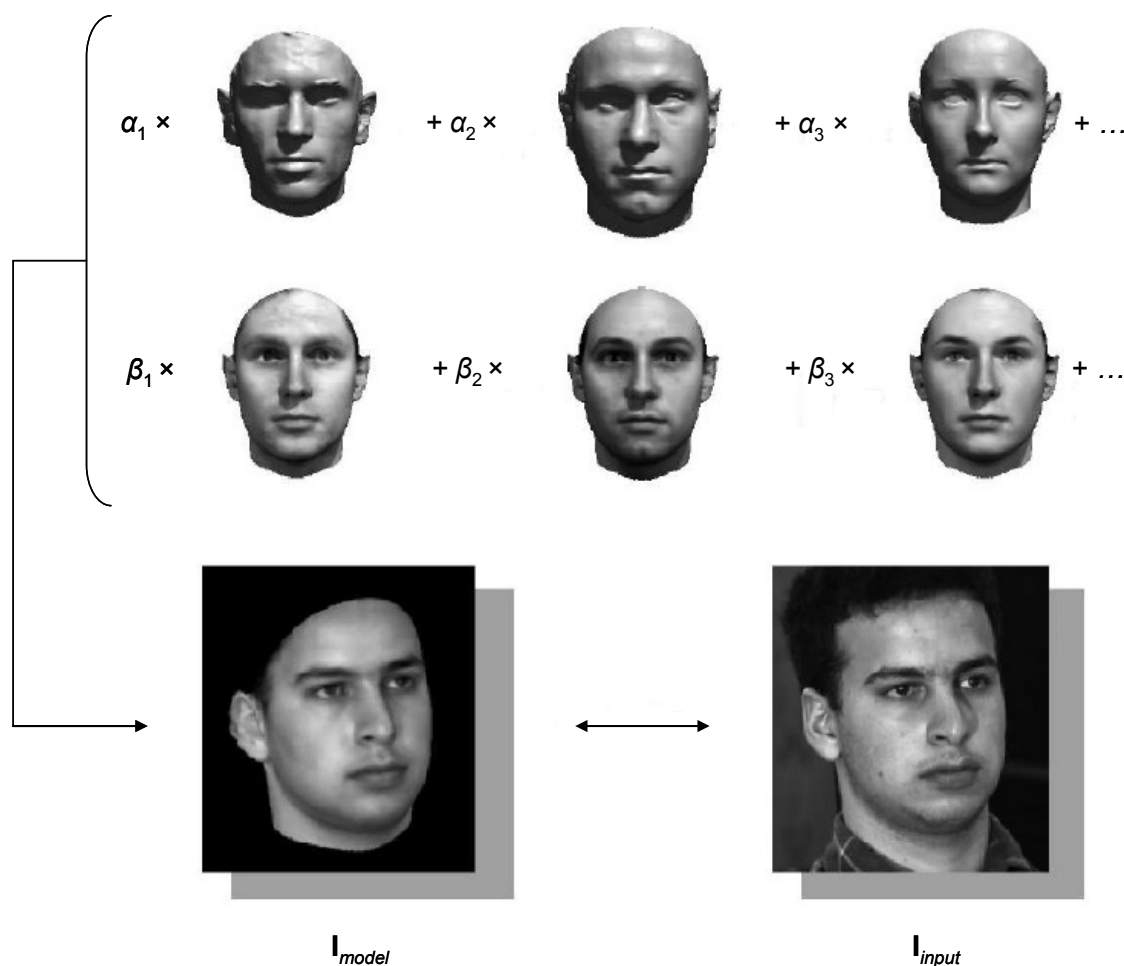
I_{model} $\longleftrightarrow$ I_{input}

Figure 8.   The fitting algorithm finds the shape coefficients α_i and texture coefficients β_i of a 3D face model such that the rendering of I_{model} is as similar as possible to I_{input}. This figure was adapted from Blanz and Vetter (2003).

the equation (16).   Appendix B provides detailed descriptions of the derivative calculations of this cost function.

In addition to optimizing the 99 shape and 99 texture coefficients[7], Blanz and Vetter (1999) implementation optimized an additional 22 rendering parameters, such as

---

[7] Their database consists of $m = 100$ faces and therefore there has 99 free parameters.
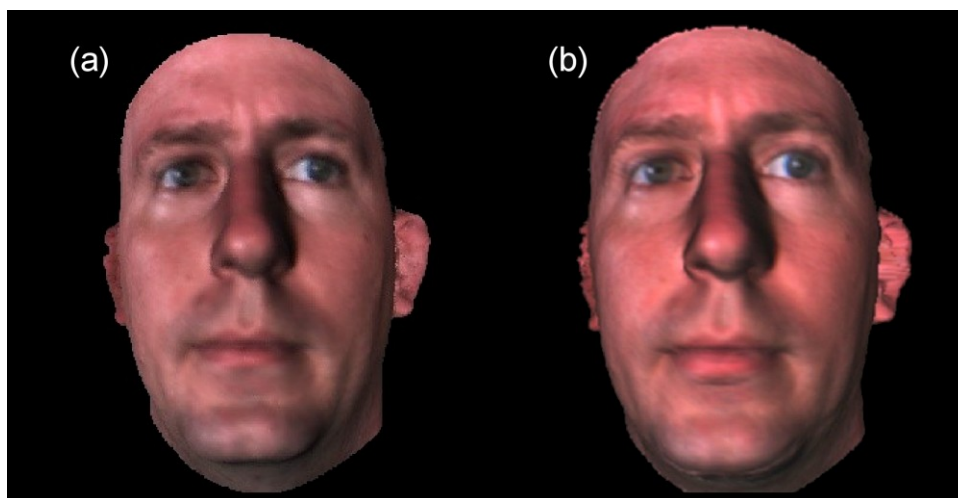
Figure 9. Reconstruction example from our morphable model implementation. (a) shows the input target image, and (b) shows the rendering of the resulting 3D model. In this instance, the 3D model of the target image was part of the 3DFS dataset.

pose angles, 3D model translation, focal length, light intensities, directional light angles,

contrast, gain and offsets of color channels. For the purposes of this research, however,

our implementation was limited to shape and texture coefficients while the rest of the

rendering parameters were assumed known. As will be argued in Section 6, this has

the advantage of decoupling the performance of the face reconstruction algorithm from

the more fundamental questions addressed in this research. Figure 9 shows a

reconstruction from our morphable model implementation. Refer to Appendix C for

more details of this implementation.

3.2 Face normalization

Distinctive faces need to be caricaturized less than typical faces in order to achieve the

same level of distinctiveness (Rhodes and McLean 1990), Therefore, to caricaturize

faces consistently and evenly, we first normalize faces so they are at the same

Mahalanobis distance[8] with respect to the norm(see Duda et al 2001) from the mean

(O'Toole et al 1997). As a result, normalized faces will lie in a hyper-ellipse (see Figure

10). Since computing this average Mahalanobis distance requires the calculation of the

full covariance matrix, the average distance between the faces and the mean will be

computed using the whitening transform instead (see Fukunaga 1990), which first

performs an orthonormal transformation (17) on A using Turk (1991):

$$Y = \Phi^T A \tag{17}$$

where $\Phi^T = (\phi_1, \phi_2, ..., \phi_n)$ are the eigenvectors of shape matrix $A = (a_1, a_2, ..., a_n)$, and

$a_i = S_i - s_{avg}$ is a mean-centered shape vector. A second transformation is then

performed to scale the principal components to unit variance:

$$Y = \Lambda^{-\frac{1}{2}} \Phi^T A \tag{18}$$

where Λ is a diagonal matrix containing the eigenvalues of the covariance of A; $\Phi^T \Lambda^{-\frac{1}{2}}$

is known as the whitening transform. Once all faces are transformed into the whitened

space, the average Euclidean distance of all faces to the origin is computed as:

_____

[8] The Mahalanobis distance is just the Euclidean distance weighted by the covariance matrix of
    the sample data.

$$d_{avg} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij}^2}{m} \tag{19}$$

where m is the number of faces and n is the number of dimensions. Faces are then rescaled to this average distance:

$$\hat{y}_i = \frac{y_i}{\|y_i\|} \cdot d_{avg} \tag{20}$$

Finally, the normalized faces $\hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m)$ are transformed back to the original space using equation (21).

$$\hat{A} = \Phi \Lambda^{\frac{1}{2}} \hat{Y} \tag{21}$$

This normalization process is illustrated in Figure 10.

3.3 Caricaturization

The caricaturization process is based on a method first proposed by Brennan (1985), where caricatures are generated by amplifying the difference between source model and a prototype or average face of a population. A caricature of a 3D face model f is defined as:

$$f_C = f_{avg} + (1 + \alpha) \cdot (f - f_{avg}); \quad \forall \alpha \in \mathbb{R} : \alpha \geq -1 \tag{22}$$
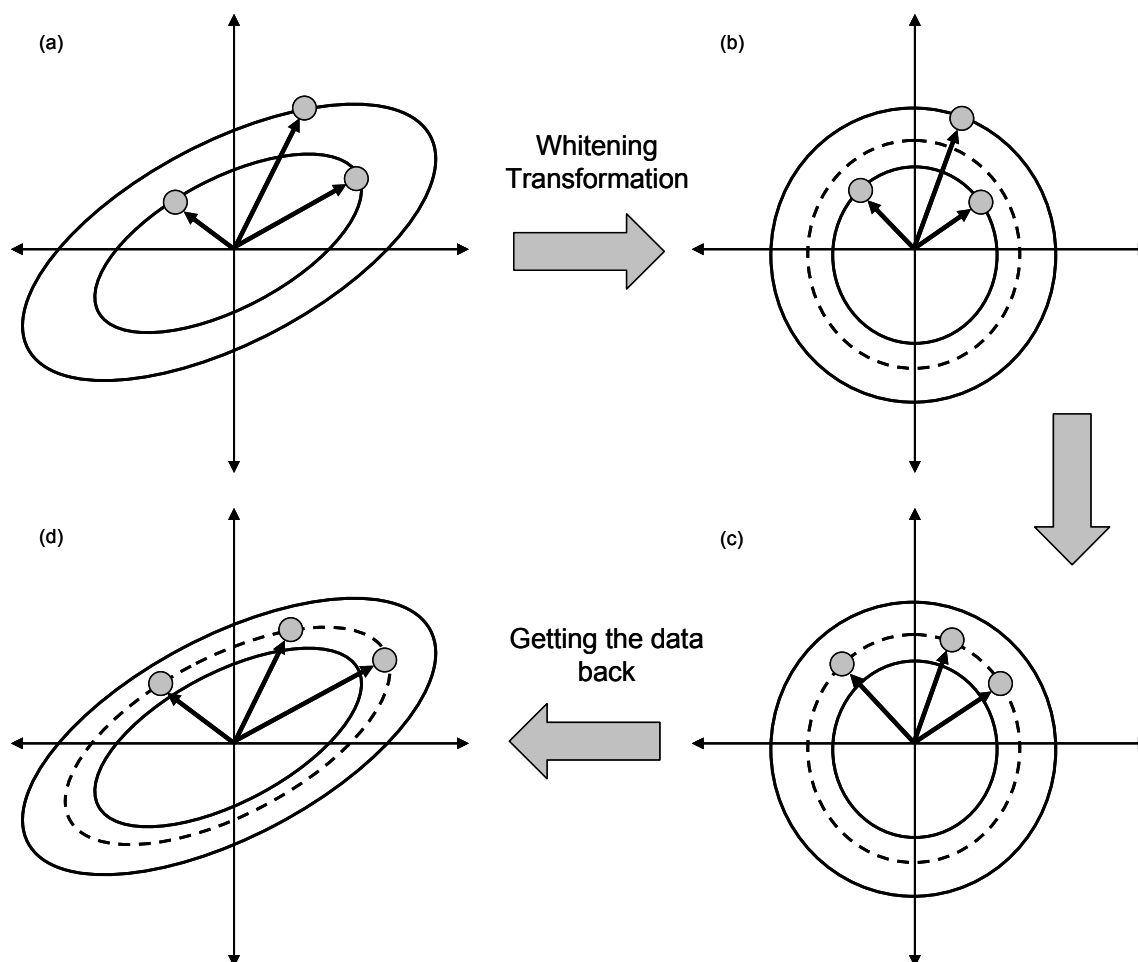
Figure 10. Face normalization. (a) Hypothetical two-dimensional face space with three faces. The origin represents the average face and the ellipses represent points at the same Mahalanobis distance from the origin. (b) The tree faces, following transformation into a whitened space. The dotted line circle represents the average Euclidean distance of all faces from the origin. (c) Faces are rescaled to be at the same average distance from the mean. (d) The original space after normalization.

where α is the caricaturization factor, f_{avg} the average face and f_C is the resulting caricature (see Figure 11). We will expand this caricaturization process later in this section, including the use of a different caricaturization factor representation that accounts for the face population distribution.
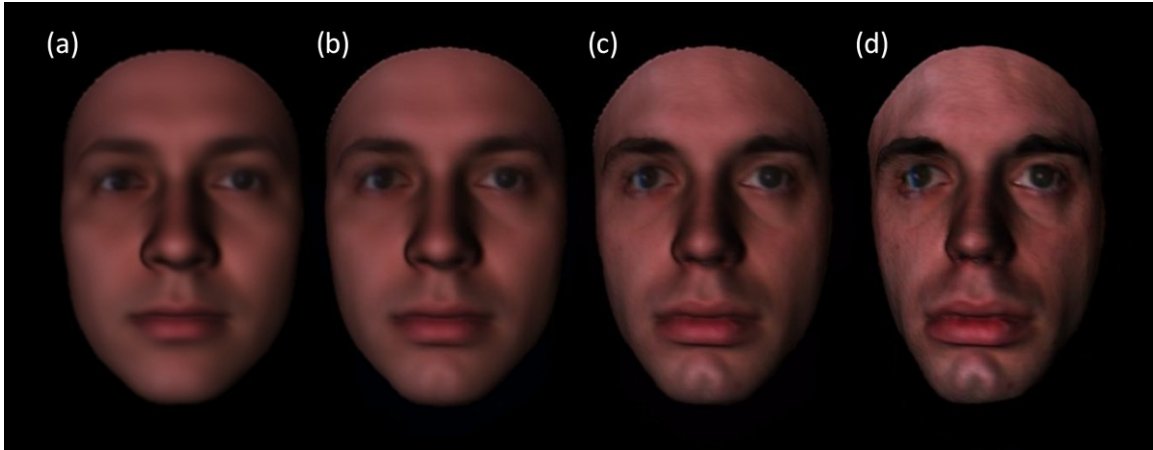
Figure 11. Average face and caricatures. The figure shows the 3DFS average face (a) f_{AVG} and the caricature (f_C) of a Caucasian male at different caricaturization levels with respect to this average face: (b) α = -0.64 or -64%, (c) α = -21%,, and (d) α = +21%.

A caricature is a transformation which exaggerates the perceptually significant features while reducing the less relevant details (Brennan 1985). Following Brennan (1985), caricatures of a 3D model are generated by amplifying the difference between the model and a prototype or average face. As indicated earlier, this process is applied to normalized faces so that a fixed caricature factor leads to the same degree of distinctiveness regardless of the face to which it is applied. A face shape S_i is defined by:

$$S_i = s_{avg} + \Delta S \tag{23}$$

where s_{avg} is the average face and ΔS is its difference from the average face shape. A linear caricature transformation is defined by:
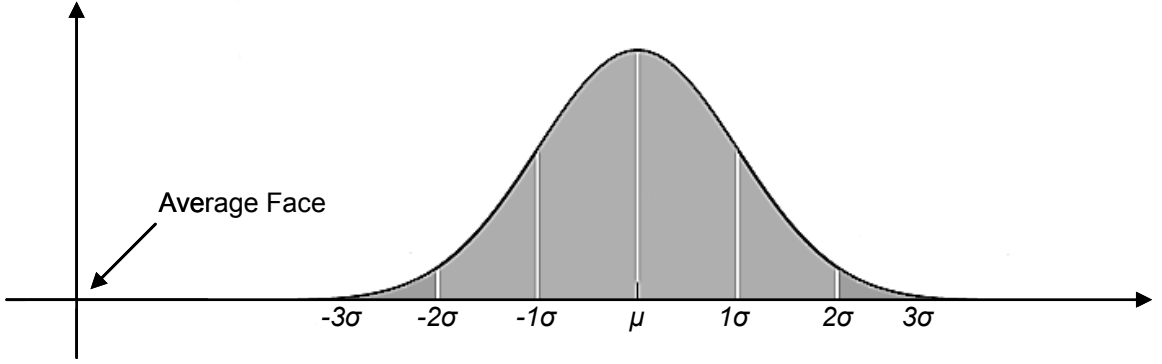
Figure 12. Exaggeration (caricature) levels spaced relative to the standard deviation (σ) of the average Euclidean distance (μ) of all faces to the mean face (origin). Caricaturization rescales a face to a level further away from the origin, whereas anti-caricaturization rescales a face closer to the origin.

$$S_i = s_{avg} + (1 + \alpha) \cdot \Delta S; \quad \forall \alpha \in \mathbb{R}: \alpha \geq -1 \tag{24}$$

where α is the exaggeration level. Alternatively, one can define the exaggeration level in terms of the standard deviation of the distance of all faces to the mean face (i.e., before normalization), illustrated in Figure 12. This allows for caricaturization levels to be based on the face population distribution and avoids over-caricaturization by keeping the levels between -3σ and $+3\sigma$. The relation between α and σ is defined by:

$$S_{avg} + (1 + \alpha)\mu = S_{avg} + (\mu + k\sigma) \tag{25}$$

$$\alpha = \frac{k\sigma}{\mu}; \quad \forall k \in \mathbb{Z} \ \text{ s.t. } \ \alpha \geq -1 \tag{26}$$

and the caricaturization equation as function of σ is defined by:
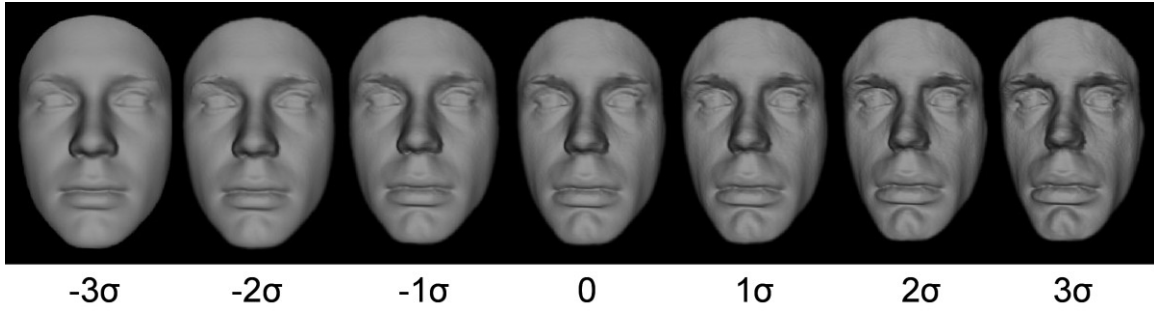
Figure 13. Rendering of a 3D face shape model at different caricaturization levels (σ). Faces are presented here without texture to highlight the anthropometric features.

$$S_i = S_{avg} + \left(1 + \frac{k\sigma}{\mu}\right)\Delta S \tag{27}$$

where μ is average Euclidean distance of all faces to the mean face. Once the 3D caricaturized face model is obtained using equation (24) or (27), the image rendering are obtained using an OpenGL face rendering tool we developed[9], which allows us to render images from 3D models based on user-configurable camera and lighting positions, and illumination intensity (see Figure 13). Figure 14 and Figure 15 show additional 3D models being caricaturized with factors ranging from +1σ to +4σ. These figures also illustrate the extent to which faces are distorted by the caricaturization process. Finally, Figure 16 and Figure 17 provide a side-by-side comparison between a set of veridical faces and their corresponding caricatures 2σ apart.

_____

[9] During the early stages of this research we used a tool developed by Hernández (2006). We developed a new rendering tool that we integrated to Matlab development environment.
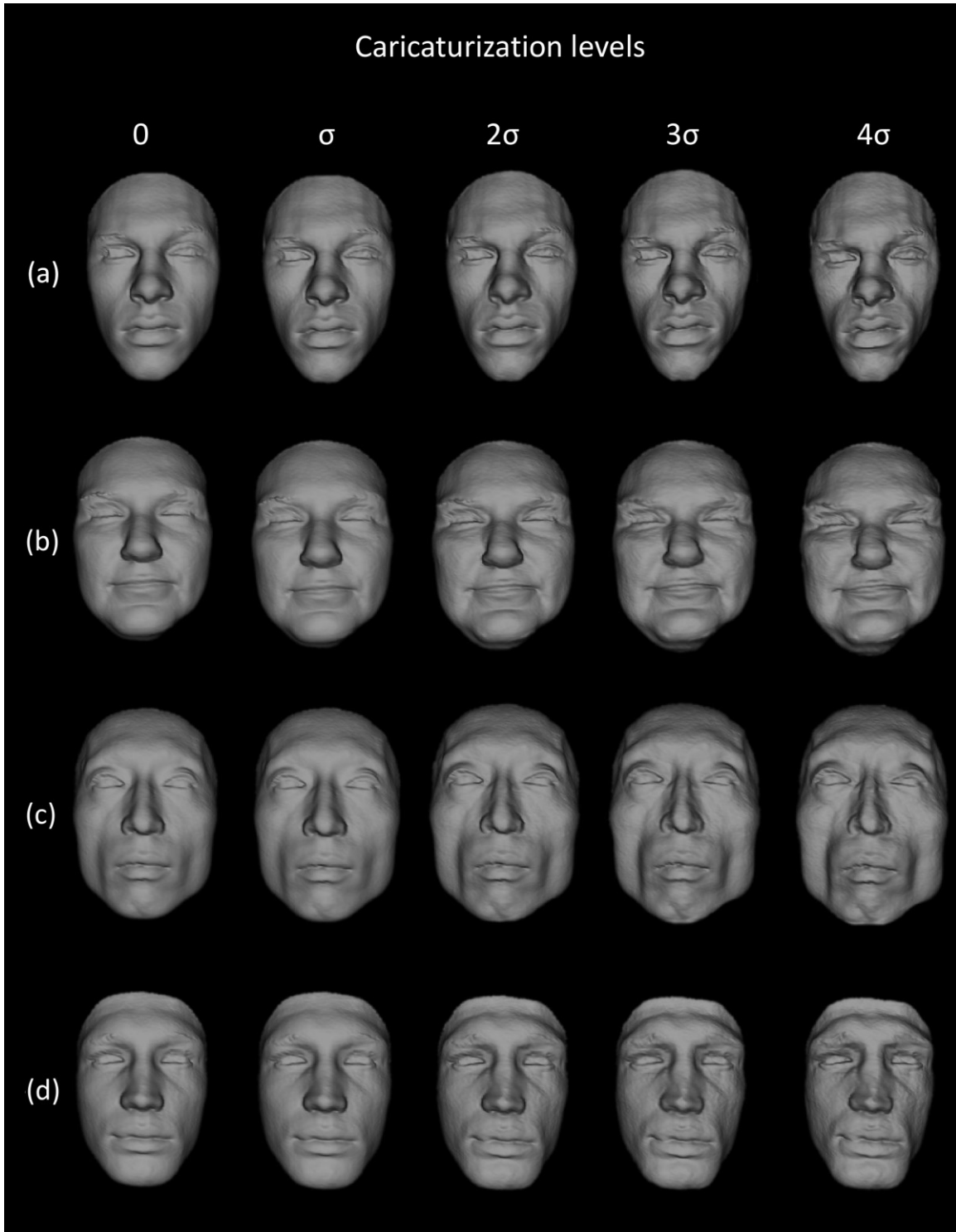
Figure 14. Shapes at different caricaturization levels. The figure shows four additional face models from the 3DFS dataset from no caricaturization (0) to high caricaturization level (+4σ).
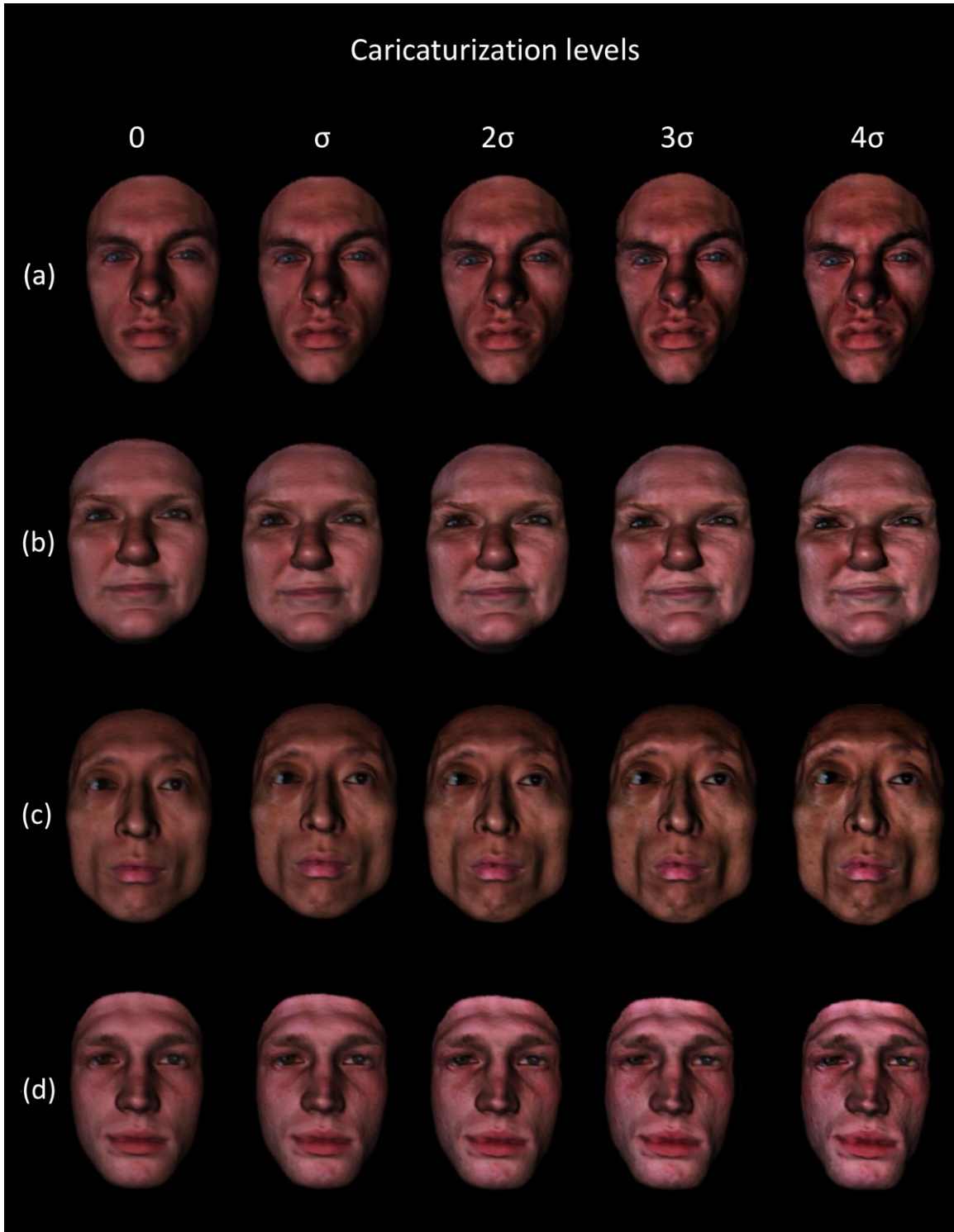
Figure 15. Shapes with textures at different caricaturization levels. The figure shows four face models from the 3DFS dataset from no caricaturization (0) to high caricaturization level (+4σ).
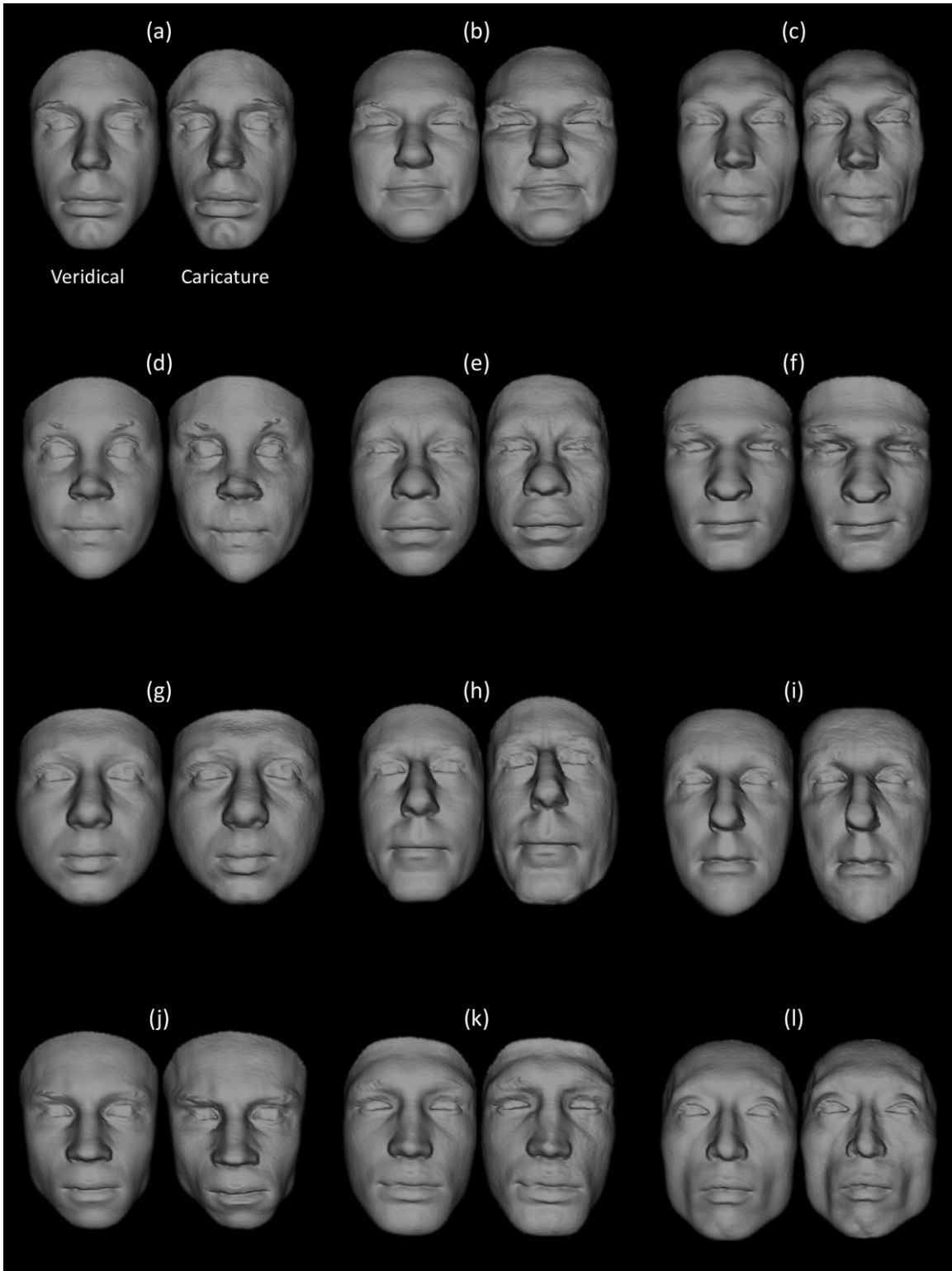
Figure 16. Pairwise caricature comparison. The figure shows several veridical 3D shape face models from the 3DFS dataset and the corresponding caricatures (2σ apart).
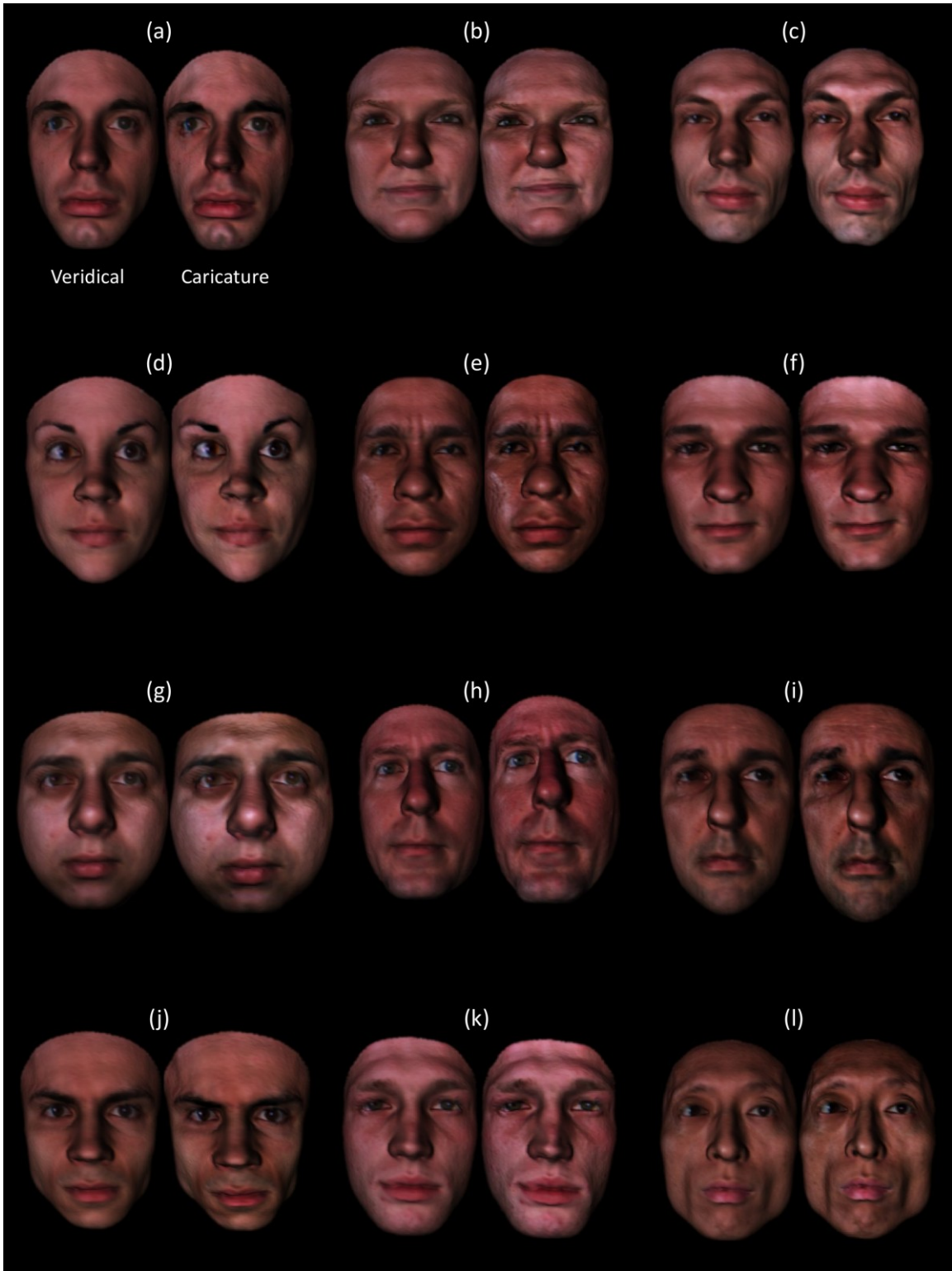
Figure 17. Pairwise caricature comparison with texture. The figure shows several veridical faces from the 3DFS dataset and the corresponding caricatures (2σ apart*)*.

4. PERCEPTUAL STUDY: CARICATURES AND REVERSE CARICATURES

The main objective of these perceptual studies that will be presented in this section, as well as in Sections 5 and 7, is to determine the extent to which the recognition of faces can be improved by enhancing their facial distinctiveness during familiarization, particularly on groups of faces that are naturally difficult to differentiate, such as members from another race or typical faces. These perceptual studies were conducted with the objective of, first, replicating some of results reported in the literature and, more importantly, validating our hypotheses.

All perceptual studies used a common paradigm for recognition tasks known as the *old/new test*. Specifically, our studies followed similar procedures to the ones outlined in Duchaine et al (2003) and Goldstein and Chance (1980). According to these procedures, subjects are first familiarized with a set target faces, each presented for a brief period in a random sequential order. Following the familiarization phase, subjects are then tested with a set of faces containing the target faces among a set of confounders. During this recognition phase, the subjects are asked to identify each face as "new" if they have not seen it before, or as "old" if they recognize it as one from the familiarization phase.

In an initial calibration study, we used a modified old/new face recognition procedure (Duchaine et al 2003; Goldstein and Chance 1980) to determine the minimal

degree of exaggeration between veridical and caricature faces that would lead to a direct caricature effect (i.e., whereby caricatures would be recognized more accurately than veridical faces) (Benson and Perrett 1994; Rhodes et al 1987). This exaggeration level was then used as the basis for each of the subsequent recognition experiments. This calibration trial also served to verify that we could replicate the distinctiveness effects reported in the literature. Then, in order to better mimic the canonical viewpoint for faces and in contrast to the procedure used by Deffenbacher et al (2000), we used the above old/new face recognition procedure to establish whether the reverse-caricature effect would manifest when only a frontal representation of unfamiliar faces was given during both training and testing (Experiment 1). In the subsequent two experiments (Experiments 2 and 3), we shortened familiarization times, and introduced random facial rotations at test. We predicted that, even under these more difficult conditions, familiarization with caricatures would produce greater recognition performance with unexaggerated faces than would familiarization with veridical versions of those faces.

4.1 Calibration study

To avoid ceiling effects, the first phase was a calibration trial to select the appropriate exaggeration levels for the veridical and caricature faces. This calibration trial also served to verify that we could replicate the distinctiveness effects reported in the literature.

4.1.1 Participants

Participants were recruited from an Introductory Psychology pool. All participants were undergraduate students (ages 18-25). Ninety-nine students ($n = 60$ female, $n = 39$ male) participated in the calibration study. A similar gender distribution for participants was maintained across experimental conditions.

4.1.2 Stimuli

A total of 40 face models were selected from the University of Freiburg 3DFS-100 dataset (Blanz and Vetter 1999). Each face consisted of a mesh with 75972 vertices, each vertex defined by its position (in 3D Cartesian coordinates) and its reflectance (in RGB space). In order to caricaturize faces consistently and evenly[10], each face f was first normalized by its Mahalanobis distance ($\| \quad \|_M$) to the average face f_{AVG} (O'Toole et al 1997) and then caricaturized by linearly exaggerating its differences with respect to f_{AVG}. The average face f_{AVG} was computed using all 100 faces from the 3DFS dataset. Seven exaggeration levels were considered: $-3\sigma, -2\sigma, -1\sigma, 0, +1\sigma, +2\sigma, +3\sigma$ where σ was defined as the standard deviation of the distance between un-normalized faces in the dataset and their average. This parameterization was chosen (as opposed to the conventional percentage factor α) because it adjusts the caricaturization level to the variability of

---

[10] Distinctive faces need to be caricaturized less than typical faces in order to achieve the same level of distinctiveness (Rhodes and McLean 1990).
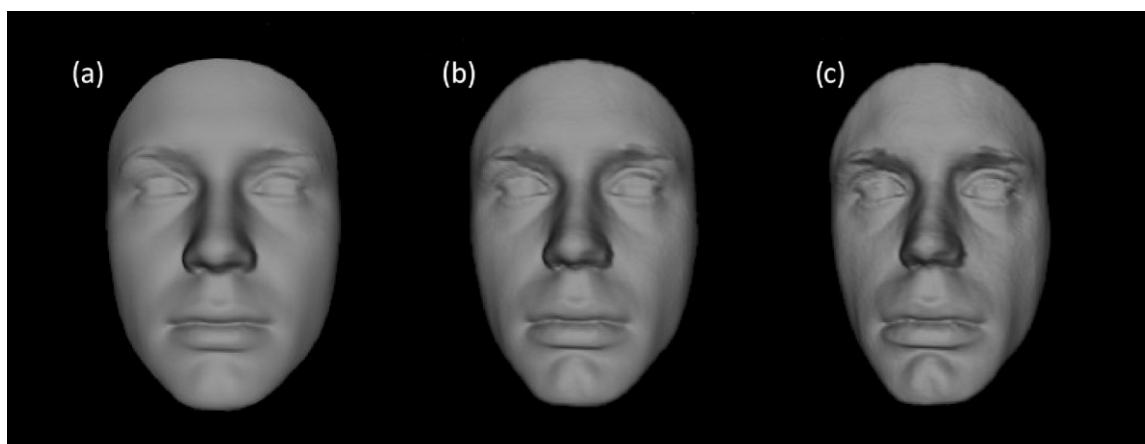
Figure 18. Sample stimuli. It shows the face of a Caucasian male without texture at different caricaturization levels: (a) -3σ, (b) -1σ, (c) $+1\sigma$. The ears and neck were removed to prevent participants from using picture-matching strategies.

faces in the dataset[11]. Refer to subsections 3.2 and 3.3 for details on face normalization

and caricature generation, respectively. As shown in Figure 18, faces were rendered

frontally without facial texture to focus attention on anthropometric features.

Presentation of stimuli was performed using the DMDX display software (Forster and

Forster 2003).

4.1.3 Procedure

Stimuli were displayed on a 17-inch monitor, configured to a 1280 by 1024 resolution at

60 Hz. Seven groups of participants (n = 13, 14, 14, 15, 15, 15, 13) were first familiarized

with 10 randomly chosen target faces. Following the experimental procedure in

---

[11] The seven exaggeration levels α = {-0.64, -0.43, -0.21, 0, +0.21, +0.43, +0.64} were selected to be separated from each other by one standard deviation, as measured on the distribution of faces in the database. Note that, for α = -1, the anti-caricature becomes the average face.

Duchaine et al (2003), each target face was presented twice for 3 seconds at a time.

Afterwards, participants were tested on 50 faces, of which 30 were new (non-target)

faces and 10 were the learned targets. The latter were presented twice to increase the

amount of data collected during calibration (i.e. each presentation was treated as a

different observation). Caricature level was held constant between familiarization and

test for each group. Target and non-target faces within a group had the same

caricaturization level. Participants were not aware of which exaggeration level they

were viewing. Order of presentation was randomized across participants for both

familiarization and test. Following Duchaine et al (2003) and Goldstein and Chance

(1980), no time limits were imposed, but participants were asked to make their decision

(i.e., old vs. new face) as rapidly as possible without sacrificing accuracy.

4.1.4 Results

In this and later experiments recognition accuracy was measured using the signal

detection's sensitivity index, or d' (Macmillan and Creelman 2005). A one-way analysis

of variance (ANOVA) on this measure revealed a main effect of condition, $F(6, 92) =$

10.45, $p < 0.001$. Face recognition performance as a function of caricaturization levels

are shown in Figure 19a. Starting from a near-chance level for extreme anti-caricatures

$\alpha = -0.64$, recognition performance increased with the level of caricaturization. Values

of $\alpha = -0.21$ (for veridical faces) and $\alpha = +0.21$ (for caricatures) were found to be
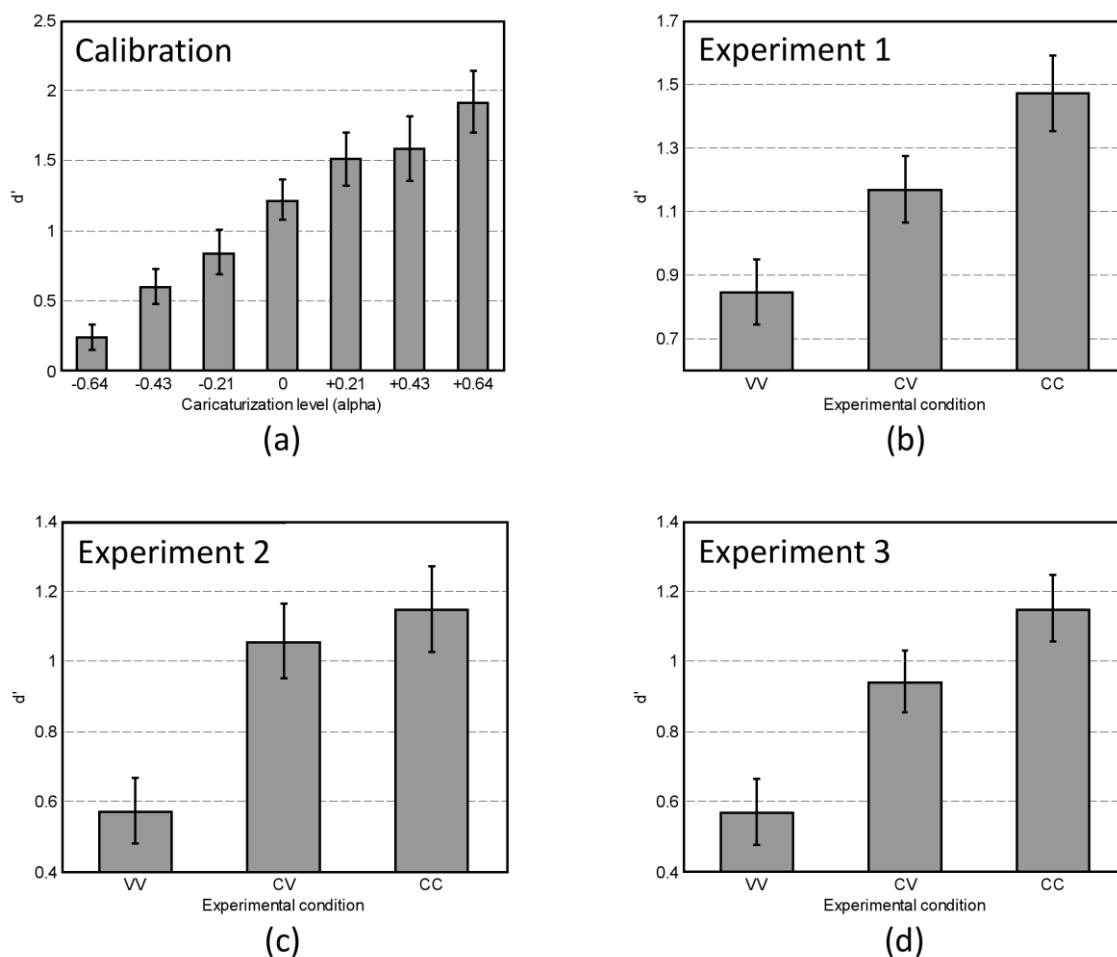
Figure 19. Face recognition accuracy on (a) the Calibration Study, (b) Experiment 1, (c) Experiment 2, and (d) Experiment 3.

the minimally effective levels of exaggeration, $t(27) = 2.55$, $p < 0.05$, and were chosen for the reverse-caricature experiments.

No significant differences were observed in response time across caricaturization levels, $F_{(6, 92)} < 1$ (one-way ANOVA). Additionally, the dual presentation of targets had no main effects on hit rates during test, $F(1, 92) < 1$, and had

Figure 20. Hit rates on the dual presentations of the targets during (a) the calibration study and (b) Experiment 1.

no interactions with the experimental condition, $F(6, 92) < 1$ (two-way mixed-factor ANOVA); hit rates on the two presentations of the targets are illustrated in Figure 20a.

4.2 Experiment 1

The goal of this experiment was to establish the reverse-caricature effect with frontal views rather than with the more favorable three-quarter views that have been used in Deffenbacher et al (2000). Additionally, participants were trained for shorter periods to test the lower limits of training times.

4.2.1 Participants

Participants were recruited from an Introductory Psychology pool. Eighty-eight students ($n = 59$ female, $n = 29$ male) participated in Experiment 1. A similar gender distribution for participants was maintained across experimental conditions.

4.2.2 Stimuli

The exaggeration levels established in the calibration study were used to select the stimuli for experiment 1 (α = -0.21 for veridical faces, and α = +0.21 for caricatures). As in the calibration study, the faces were rendered without the reflectance information.

4.2.3 Procedure

The procedures were the same as in the calibration study, but participants were assigned to one of three experimental conditions: (*i*) veridical familiarization/veridical test (*n* = 30), (*ii*) caricature familiarization/caricature test (*n* = 29), and (*iii*) caricature familiarization/veridical test (*n* = 29). In all three conditions (VV, CC, or CV) the same ten randomly chosen faces were used (in either caricature or veridical form). The VV condition established the experimental baseline while the CC condition established an upper bound on facial recognition performance. The CV condition tested our working hypothesis, according to which familiarization with caricatures improves the recognition performance when presenting faces in their veridical form at a later time.

4.2.4 Results

A one-way ANOVA on d' revealed a main effect of condition on recognition accuracy, $F(2, 85)$ = 7.96, p < 0.001. Face recognition performance of the three experimental conditions is shown in Figure 19b. Consistent with other researchers' findings and our own calibration study, participants exposed to caricature targets and caricature probes (CC) showed higher recognition accuracy than participants shown veridical targets and

veridical probes (VV), $t(57) = 3.90$, $p < 0.001$. Of greater interest for the current study, participants familiarized with caricature targets and tested on veridical probes (CV) had higher overall recognition accuracy than those shown veridical targets and veridical probes, $t(57) = 2.17$, $p < 0.05$. Looking at the experimental condition and the participants' gender as factors, we found no main effect of gender $F(1,82) = 0.052$, *ns*. Namely, males and females had similar recognition performance.

No significant differences were observed in response time across the three conditions, $F(2, 85) < 1$ (one-way ANOVA). Additionally, the dual presentation of targets had no main effects on hit rates dur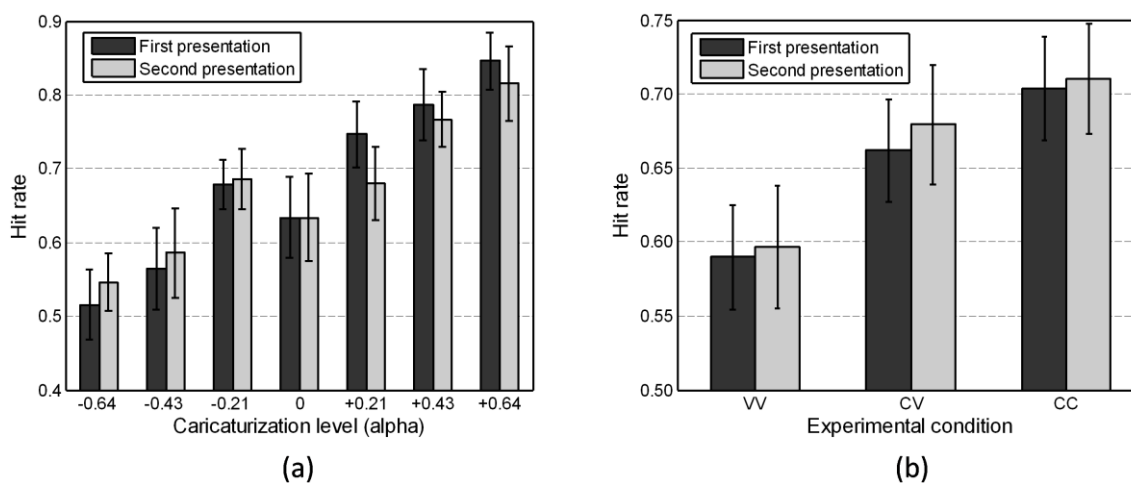ing test, $F(1, 85) < 1$, and had no interactions with the experimental condition, $F(2, 85) < 1$ (two-way mixed-factor ANOVA); hit rates on the two presentations of the targets are illustrated in Figure 20(b).

4.3 Experiment 2

The goal of this experiment was to establish whether a reverse-caricature effect would occur given a shorter familiarization time than was used in Experiment 1.

4.3.1 Participants

Participants were recruited from an Introductory Psychology pool. One hundred and two students ($n = 74$ female, $n = 28$ male) participated in Experiment 2. A similar gender distribution for participants was maintained across experimental conditions.

4.3.2 Stimuli

The faces used in this experiment were the same as in Experiment 1.

4.3.3 Procedure

Participants were again assigned to one of three experimental conditions: VV (n =34), CC (n = 34), and CV (n = 34), where they were familiarized with 10 randomly selected faces. However, unlike in the previous experiments, target faces were presented only once for 3 seconds (see Goldstein and Chance 1980). Participants were then tested with a single presentation of 40 faces, of which 30 were new (non-target) faces and 10 were the learned faces. These procedures were introduced to increase the difficulty of the recognition test. In all three conditions, the same ten randomly chosen faces were used (in either caricature or veridical form). The exaggeration levels established in the calibration study were used for all faces (α = -0.21 for veridical faces, and α = +0.21 for caricatures).

4.3.4 Results

A one-way ANOVA revealed a main effect for recognition accuracy (d') across the three groups, $F(2, 99) = 7.78$, $p < 0.001$. Face recognition performance on the three experimental conditions is shown in Figure 19c. Participants exposed to caricature targets and caricature probes (CC) showed higher recognition accuracy than participants exposed to veridical targets and veridical probes (VV), $t(66) = 3.65$, $p < 0.001$. More importantly, participants familiarized with caricature targets and tested on

veridical probes (CV) had higher overall recognition accuracy than those shown

veridical targets and veridical probes (VV), $t(66) = 3.31$, $p < 0.01$. No significant

differences were observed in response time across the three conditions, $F(2, 99) < 1$ (one-

way ANOVA). Looking at the experimental condition and the participants' gender as

factors, we found no main effect of gender $F(1,96) = 2.81$, *ns*. Namely, males and

females had similar recognition performance.

4.4 Experiment 3

The goal of Experiment 3 was to rule out the possibility that the results of the previous

two experiments were based on the use of a superficial "picture matching" strategy by

participants rather on actual face recognition. To this end, test faces were rotated

relative to their frontal orientation during familiarization.

4.4.1 Participants

Participants were recruited from an Introductory Psychology pool at Texas A&M

University. One hundred and two students ($n = 76$ female, $n = 26$ male) participated in

Experiment 3. A similar gender distribution for participants was maintained across

experimental conditions.
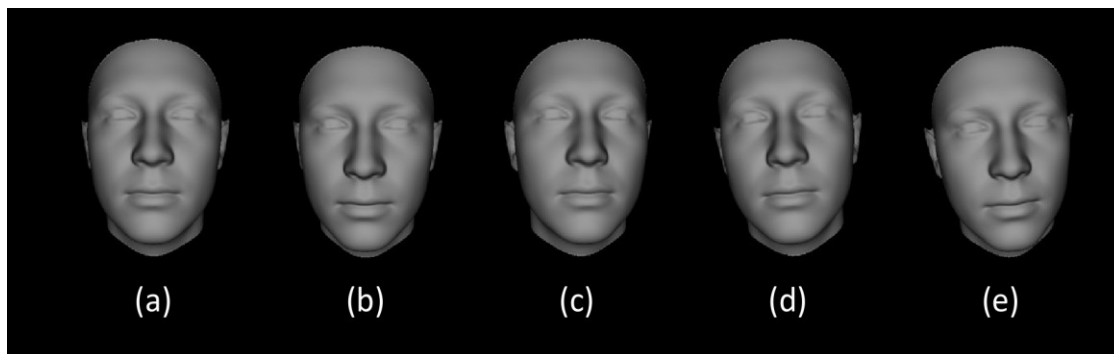
Figure 21. Example of face rotations. (a) shows a 3D model with no rotation. (b) shows the model rotated 5 degrees on the X axis, (c) 5 degrees on the Y axis, and (d) 5 degrees on the Z axis. Finally, (e) shows 5 degrees on all axes combined.

4.4.2 Stimuli

The same stimulus set from Experiment 1 and 2 were used in this study. However, the test faces were rendered with a uniform distributed random rotations in the range of ±5 degrees in each of the three axes. An example of these rotations is shown in Figure 21.

4.4.3 Procedure

Participants were again assigned to one of three experimental conditions: VV (n =34), CC (n = 34), and CV (n = 34), using identical procedures to the previous studies, with two exceptions. First, participants were familiarized with a single presentation of the target faces for 6 seconds each. Second, all faces were presented with a modest random rotation during test.

4.4.4 Results

A main effect was observed for recognition accuracy (d') across the three groups, $F(2,$ 99) = 9.46, $p < 0.001$. Face recognition performance on the three experimental conditions is shown in Figure 19d. Participants exposed to caricature targets and caricature probes (CC) showed higher recognition accuracy than participants shown veridical targets and veridical probes (VV), $t(66) = 4.21$, $p < 0.001$. More importantly, participants familiarized with caricature targets and tested on veridical probes (CV) had higher overall recognition accuracy than those shown veridical targets and veridical probes (VV), $t(66) = 2.79$, $p < 0.01$. No significant differences were observed in response time across the three conditions, $F(2, 99) < 1$. Looking at the experimental condition and the participants' gender as factors, we found a main effect of gender $F(1,96) = 5.13$, $p < 0.05$. Females performed better than males overall, $t(100) = 1.53$, ns ($p = 0.13$).

4.5 Summary

Results from the three reverse-caricature studies support the hypothesis that even very brief familiarization with frontally presented caricatures leads to greater recognition of their unexaggerated counterparts than does familiarization with veridical versions of those faces. Taking the results from Experiment 1 as a baseline, Experiment 2 explored the effect of shorter familiarization times (3-sec vs. 6-sec) on recognition performance, whereas Experiment 3 explored the effect of random rotations at test (±5 degrees); both Experiments 2 and 3 also explored the effect of single presentations at test. Our results

indicate that the reverse-caricature effect is robust to differences in stimuli presentation

schemes, familiarization times, and modest rotations of the faces.

5. PERCEPTUAL STUDY: REVERSIBILITY OF OTHER-RACE

EFFECTS

In this study we investigate whether this reverse-caricature effect can be exploited to reduce other-race effects, i.e. by familiarizing subjects with other-race faces whose distinctiveness has been purposely enhanced. To achieve this objective, we included texture in our 3D face models. Using the automated caricature generation method and 3D facial models described in Section 3, we tested our working hypothesis through a series of perceptual experiments focused on own-race vs. other-race recognition.

We employed a similar old/new face recognition protocol described in Section 4, where subjects are familiarized with a set of faces, and then asked to recognize those faces among a set of confounders. In an initial calibration study, we used this method to determine (*i*) the minimal degree of exaggeration between veridical and caricature faces that would lead to a caricature effect, and (*ii*) the level and races that would likely produce a significant other-race effect *on this specific dataset*. This exaggeration level was then used as the basis for a recognition study that evaluated whether a brief initial exposure to frontally-presented caricaturized faces would reduce the other-race effect when tested using their veridical counterparts.

5.1 Calibration study

To avoid ceiling effects, the first phase was a calibration trial to select the appropriate exaggeration levels for the veridical and caricature faces. This additional calibration

step was deemed necessary given that the faces were rendered with shape *and* texture, whereas the study reported in Section 4 had used only shape. Thus, given the additional information provided by facial texture, it was possible that suitable caricaturization levels would differ from those obtained on shape-only faces.

5.1.1 Participants

A total of 53 undergraduate Caucasian students (38 females and 15 males) from the Department of Psychology at Texas A&M University participated in the calibration study.

5.1.2 Stimuli

A total of 40 face models were selected from the University of Freiburg 3DFS-100 dataset (Blanz and Vetter 1999). Each face consisted of a mesh with 75972 vertices, each vertex defined by its position (in 3D Cartesian coordinates) and its reflectance (in RGB space). Following Furl et al (2002), the race distribution for these models was 12 Caucasian, 12 East Asian, 8 Indians, 4 Africans, 4 Other (Middle Eastern, Hispanic, or any other group that did not fit in the first 4 groups). Indian, African, and Other faces were considered as filler stimuli since we had initially selected East Asians as the other-race target for our pool of Caucasian subjects.

In order to caricaturize faces consistently and evenly, each face f was first normalized by its Mahalanobis distance ($|| \ ||_M$) to the average face f_{AVG} (O'Toole et al 1997) and then caricaturized by linearly exaggerating its differences with respect to f_{AVG}
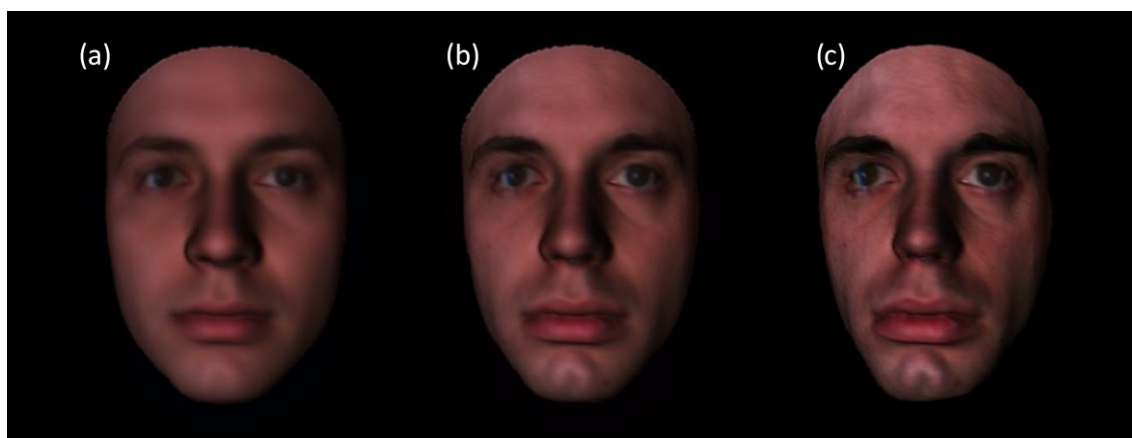
Figure 22. Sample stimuli with texture. It shows the face of a Caucasian male at different caricaturization levels: (a) -3σ, (b) -1σ, (c) +1σ. The ears and neck were removed to prevent participants from using picture-matching strategies.

(see subsections 3.2 and 3.3.) The average face f_{AVG} was computed using all 100 faces from the 3DFS dataset. Five exaggeration levels were considered: -3σ, -2σ, -1σ, 0, +1σ, where σ was defined as the standard deviation of the distance between un-normalized faces in the dataset and their average. Figure 22 shows an example of the stimuli used in this experiment.

Renderings of the resulting 3D caricatures for three caricaturization levels and average face are shown in Figure 22. Presentation of stimuli was performed using the DMDX display software (Forster and Forster 2003).

5.1.3 Procedure

Participants were first assigned to one of five experimental conditions, maintaining a similar gender distribution across condition (n=10, 11, 10, 11, 11). Each condition

consisted of stimuli generated at a given caricaturization level (from -3σ to +1σ).

Participants were familiarized with 20 frontal target faces, each presented twice in

random order, 3 seconds per presentation. After the familiarization phase was

completed, participants were tested on 40 faces, of which 20 faces were "new" (non-

target) and 20 faces were "old" (target); all faces in the test phase were rendered with a

random orientation between ±5 degrees in the three axes. Participants were asked to

identify each face as "old" if they recognized it as one from the familiarization phase, or

as "new" if they had not seen it before. Following (Duchaine et al 2003), no time limits

were imposed, but participants were asked to make the identification as rapidly as

possible without sacrificing accuracy.

5.1.4 Results

Significant differences were observed for recognition accuracy across the five

caricaturization levels, $F(4,48) = 20.17$, $p < 0.001$. Figure 23a shows the recognition

performance across races, excluding African and Other faces, which were significantly

under-represented in the stimulus set[12]. Starting from a near-chance level ($d' \approx 0.0$) for

extreme anti-caricatures (-3σ), recognition accuracy increased with the level of

caricaturization. Additionally, Figure 23a shows other-race effects on both East Asian

and Indian faces, but the effect is stronger on Indian faces. From these results,

_____

[12] As a result, we found that faces from these two races could be singled out and recognized
quite easily.
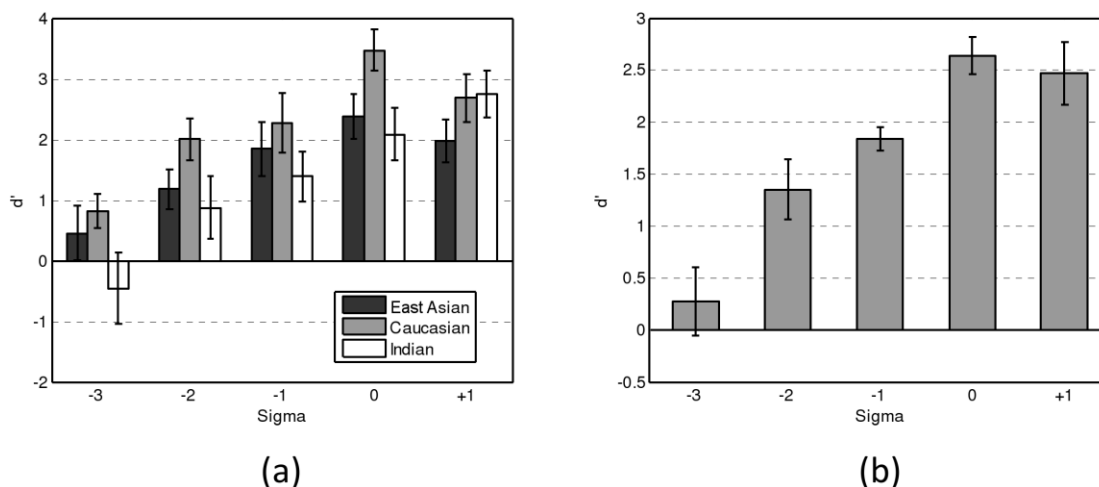
Figure 23. Calibration study recognition results. (a) Recognition accuracy for the calibration studies across races. (b) Same results as in (a) but combining the East Asian, Caucasian, and Indian results. Error bars represent standard errors.

exaggeration levels of -1σ and 0σ were selected for veridical and caricaturized faces, respectively, on the basis that they were the closest pair with the largest statistically-significant difference in recognition accuracy (see Figure 23b): $M_{0\sigma} = 2.64 \pm 0.18$ (Mean $\pm$ S.E.M.), $M_{-1\sigma} = 1.85 \pm 0.12$, $t(19) = 3.73$, $p<0.01$ (two-tailed[13]); and they were far from the upper and lower saturation levels, which avoided potential ceiling effects. Furthermore, these caricaturization levels showed a other-race effect on the Indian stimuli for our selected veridical faces: $M_{Caucasian(-1\sigma)} = 2.28 \pm 0.49$, $M_{Indian(-1\sigma)} = 1.38 \pm 0.41$, $t(18) = 1.76$, $p = 0.09$, *ns*. Response times are shown in Figure 24; significant differences were found across caricaturization levels, $F(4,48) = 4.61$, $p < 0.01$, but not within caricaturization levels.

_____

[13] All *t*-test are two-tailed unless indicated otherwise.
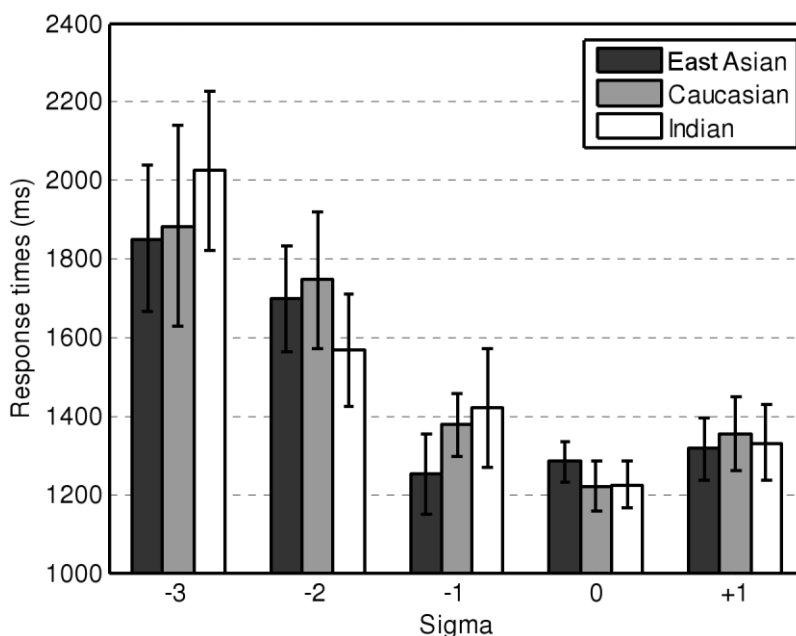
Figure 24. Response time across races. It shows that participants responded faster as we increased the caricaturization level. Error bars represent standard errors.

An interesting result from this calibration experiment is that other-race effects were stronger for Indian faces than for East Asian faces. Namely, the other-race effect was significant for Indian faces at all caricaturization levels except for $\sigma = +1$, whereas significant effects for East Asian faces were only observed for $\sigma = -2$ and $\sigma = 0$. While we do not have a precise explanation for this result, our conjecture is that either the specific Indian faces in our dataset were more homogeneous than the East Asian faces, or our participants had more exposure to East Asian faces, either socially or through the media.

5.2 Reversibility study

Having identified suitable exaggeration levels that yielded caricature effects and other-race effects, a second study was performed to test our working hypothesis: namely, that reverse-caricature training can reduce other-race effects.

5.2.1 Participants

A total of 45 undergraduate students (24 females and 21 males) from the Department of Psychology at Texas A&M University participated in this study.

5.2.2 Stimuli

Based on our calibration study, we used exaggeration levels of -1σ and 0σ for veridical and caricaturized faces, respectively. Given that the other-race effect was found to be stronger on Indian faces during the calibration study, we adjusted the stimulus dataset to balance the distribution of faces: 10 East Asian, 10 Caucasian, 10 Indian, 6 African, and 4 Other. A similar gender distribution for participants was also maintained across experimental conditions.

5.2.3 Procedure

The procedures were the same as in the calibration study, but participants were assigned to one of two experimental conditions: (VV) veridical familiarization-veridical test ($n = 23$), which served as a control group, and (CV) caricature familiarization-veridical test ($n = 22$), which tested our hypothesis that reverse-caricature training can
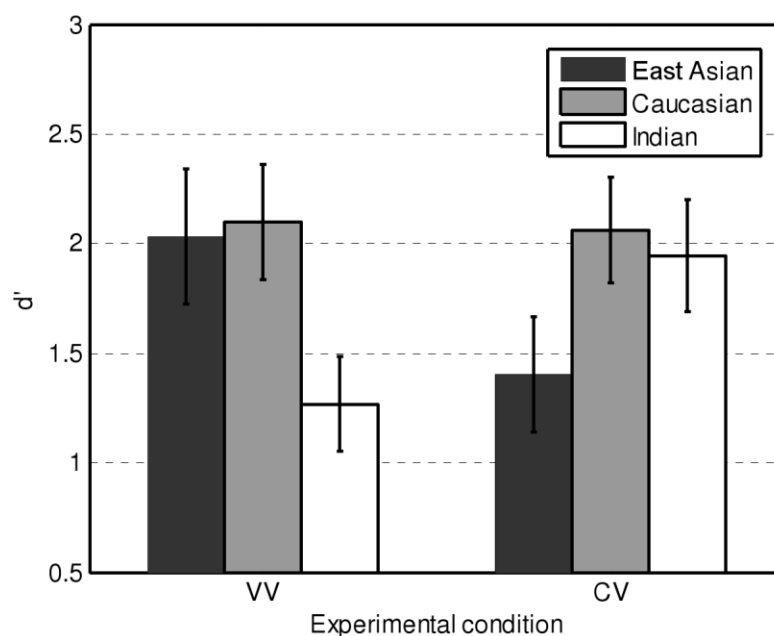
Figure 25. Recognition accuracy (*d′*) across races. Participants in the VV condition were familiarized with veridical faces and then tested with veridical faces. Participants in the CV condition were familiarized with caricatures, but tested with veridical faces. Error bars represent standard errors.

reduce other-race effects. In all conditions, the same twenty randomly chosen faces were used (in either caricature or veridical form).

5.2.4 Results

Figure 25 summarizes the results of this experiment. Participants in the VV condition showed a significant other-race effect on Indian faces: $M_{Caucasian} = 2.10 \pm 0.26$, $M_{Indian} = 1.27 \pm 0.22$, $t(44) = 2.43$, $p < 0.05$. Participants in the CV condition, on the other hand, did not show significant other-race effects on Indian faces: $M_{Caucasian(CV)} = 2.06 \pm 0.24$, $M_{Indian(CV)} = 1.94 \pm 0.25$, $t(42) = 0.33$, *ns*. Moreover, participants in the CV condition showed a significant improvement on Indian faces when compared to participants in the VV

condition: $M_{Indian(CV)}$ = 1.94 ± 0.25, $M_{Indian(VV)}$ = 1.27 ± 0.22, $t(43)$ = 2.03, $p < 0.05$. These results suggest that the previously reported reverse-caricature effect (Deffenbacher et al 2000; Stevenage 1995) can also help reduce difficulties recognizing other-race faces.

In this study we did not find a significant improvement in recognition performance of Caucasian (own-race) faces. Participants had almost no change in recognition accuracy between the CV and VV conditions: $M_{Caucasian(CV)}$ = 2.06 ± 0.24, $M_{Caucasian(VV)}$ = 2.10 ± 0.26, $t(43)$ = 0.10, ns. This result is consistent with the fact that our participants were experts[14] in identifying Caucasian faces; thus, it is to be expected that caricaturization would not provide as much benefit with Caucasian faces as it would with Indian faces. However, the inclusion of texture in the stimuli may have provided enough information to render our Caucasian faces unique. Caucasians often use skin tones and eye color when describing faces (Ellis et al 1975). Looking at participants' gender, experimental condition, and stimulus race as factor, we found a main effect of gender, $F(1,41)$ = 4.44, $p < 0.05$. Namely, the recognition performance between females and males was significantly different. For instance, on the VV condition, females did not show a significant other-race effect on Indian faces, $t(11)$ = 1.16, ns; while males had

_____

[14] By "expert" we imply that the participant's attention is already directed towards the most diagnostic features for that race. Therefore, caricaturization of those features would not represent a significant advantage. This suggests that a more difficult experiment would be required to show a reverse-caricature effect on own-race faces. This has been the case of previous reverse-caricature reports which, unlike in our study, rendered faces without surface reflectance, and instead employed either line drawings or shape-only 3D models (Deffenbacher et al 2000; Stevenage 1995).

a significant other-race effect, $t(10) = 2.56$, $p < 0.05$. Finally, consistent with our calibration results, participants did not show a significant other-race effect on East Asian faces. As we argued in the calibration study, this may be the result of our participants having had more extensive prior exposure to East Asian faces. Figure 25 shows that participant in the CV condition actually experienced a decline in performance on East Asian faces when compared to participants in the VV condition. Although this reduction was not statistically significant, our expectation was that performance on East Asian faces would have followed that on Caucasian faces. Further experimentation will be required to identify the source of this discrepancy.

5.3 Summary

Results from the reversibility study support the hypothesis that reverse-caricatures are able to reduce other-race effects. Specifically, our Caucasian participants showed other-race effects on Indian faces, which were reduced after familiarization with caricatures of those faces. However, Caucasian participants did not show other-race effects on East Asian faces. Our conjecture is that our Caucasian participants had a higher degree of prior exposure with East Asian faces than with Indian faces, thru popular culture. It is also possible that, within the 3DFS dataset, the *specific* East Asian face samples may be more perceptually distinctive than Caucasian or Indian faces. The 3DFS only has 100 face models and it is not race balanced.

Our finding that reverse-caricature training can reduce other-race effects suggests that this may be an effective paradigm to help individuals focus their attention on features that are useful for recognition. These results have important implications for the development of training tools for law enforcement and security.

6. RECONSTRUCTION OF 3D FACE MODELS

In Sections 4 and 5 we described perceptual experiments that tested the feasibility of using reverse caricatures for improving face recognition. These experiments were performed using faces for which 3D facial models were available. As discussed earlier, however, 3D facial models are expensive and in many cases not available (as when the subject is at large). For this reason, this dissertation explores whether 3D reconstructions from 2D photographs may be used instead of ground-truth 3D models. This section describes the 3D reconstruction methods that we have used for this purpose, whereas Section 7 describes perceptual results when 3D reconstructions (rather than ground-truth 3D models) are used for reverse caricaturization.

Our face reconstruction method is based on the morphable model (Blanz and Vetter 1999), a technique capable of reconstructing photorealistic 3D models of a human face from single images. The morphable model represents faces as linear combinations of prototype faces using principal components analysis (PCA). Given a 2D input image, the technique proposed by Blanz and Vetter uses an iterative non-linear optimization method to estimate the principal components and rendering parameters (e.g. illumination, camera parameters) of a 3D model such that, when rendered, is as close as possible to the 2D image. For the purposes of this research, however, our implementation was limited to shape and texture coefficients while the rest of the rendering parameters were assumed known. This was advantageous because the

reconstruction problem is linear in the shape and texture parameters, for which an optimal solution can be found through least squares. This allowed us to isolate reconstruction errors introduced by the optimization technique from those that are due to the (lack of) diversity in the training datasets.

6.1 Holistic reconstructions

For these reconstructions, we used the University of Freiburg 3DFS-100 dataset (Blanz and Vetter 1999) containing one hundred faces. Each face consisted of a mesh with $n =$ 75,972 vertices, and each vertex was defined by its position (in 3D Cartesian coordinates $S = (X_1, Y_1, Z_1, X_2, ..., Y_n Z_n)$ and its reflectance in RGB space $T = (R_1, G_1, B_1, R_2, ..., G_n B_n)$.

Because any reconstructed face must be a linear combination of the prototypes, and we assumed several rendering parameters fixed, we replaced the non-linear optimization stage in Blanz and Vetter (1999) with a PCA back-projection, which represents a best-case reconstruction in the least-squares sense.

To illustrate the performance of the PCA back-projection algorithm, we performed a leave-one-out reconstruction experiment. Namely, for each face in the 3DFS-100 dataset (i.e. the test face $f_{test} = \{S_{test}, T_{test}\}$), we first remove that face from the dataset, and then perform a PCA decomposition of the remaining $m = 99$ training faces according to their shape (S_{train}) and reflectance (T_{train}):

$$S_{train} = s_{avg} + \sum_{i}^{m-1} \alpha_i \cdot s_i \tag{28}$$

$$T_{train} = t_{avg} + \sum_{i}^{m-1} \beta_i \cdot t_i \tag{29}$$

where s_i and t_i denote the shape and reflectance eigenvectors, s_{avg} and t_{avg} are the average shape and reflectance, and α_i and β_i are the principal components of a particular face, respectively.

Next, we project the test face f_{test} along the PCA eigenvectors of the training data to obtain the predicted principal components α_{test} and β_{test} :

$$\alpha_{test} = \left(S_{test} - s_{avg}\right)^{\intercal} \cdot s_i \tag{30}$$

$$\beta_{test} = \left(T_{test} - t_{avg}\right)^{\intercal} \cdot t_i \tag{31}$$

from which a reconstructed 3D model f'_{test} is finally obtained:

$$S'_{test} = s_{avg} + \sum_{i}^{m-1} \alpha_{test} \cdot s_i \tag{32}$$

$$T'_{test} = t_{avg} + \sum_{i}^{m-1} \beta_{test} \cdot t_i \tag{33}$$

The predicted α_{test} and β_{test} coefficients represent the best solution that could be obtained with the non-linear optimization in (Blanz and Vetter 1999).
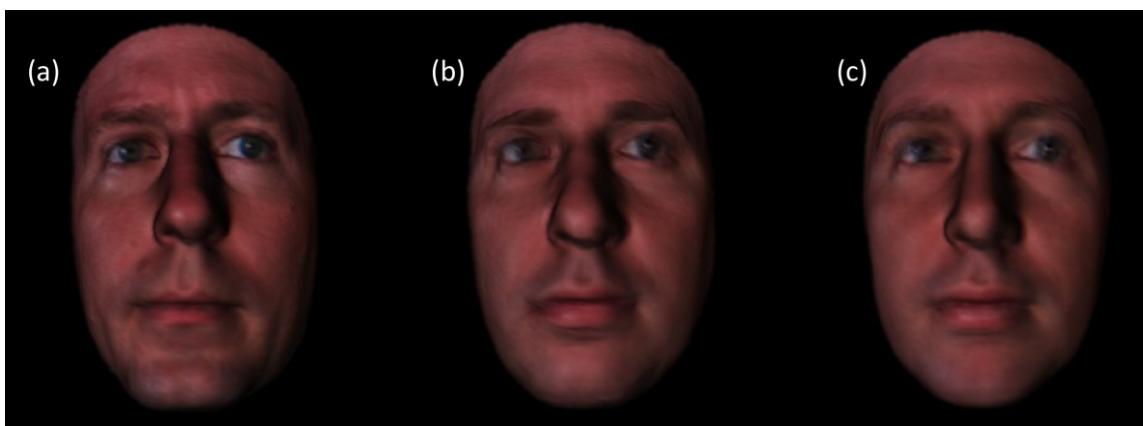
Figure 26. Comparison of linear vs. non-linear reconstruction algorithm. (a) shows a ground-truth 3D face model. (b) shows the holistic PCA back projection (least squares) and (c) shows reconstruction with the iterative algorithm in Blanz and Vetter (1999). (b) has perceptually better likeness to (a) than (c). Reconstructions done in a LOO fashion.
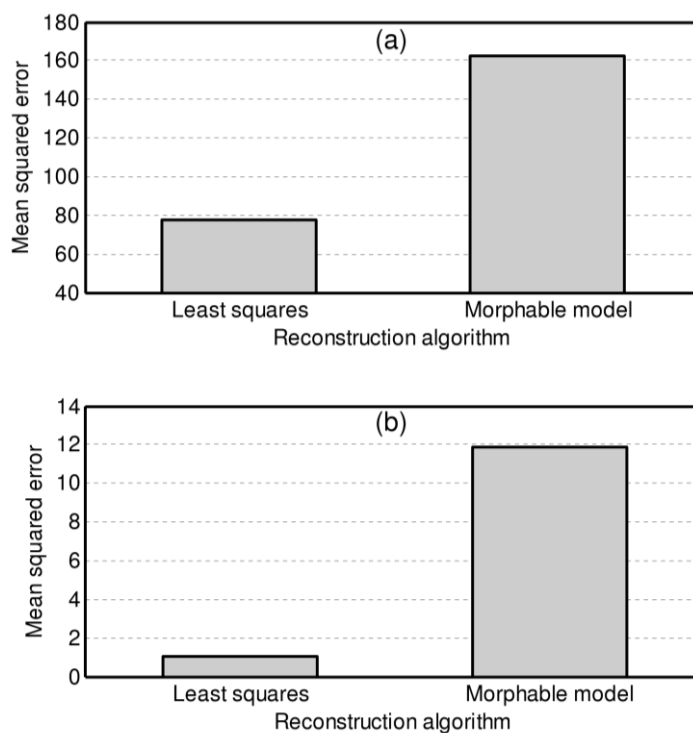
Figure 27. Error based on two reconstruction algorithms. Reconstruction based on least squares (PCA) had smaller error than using the stochastic Newton method in the morphable model in both (a) texture and (b) shape. In our implementation rendering parameters were assume fixed and equal in both algorithms, the least squares solution is the optimum solution.
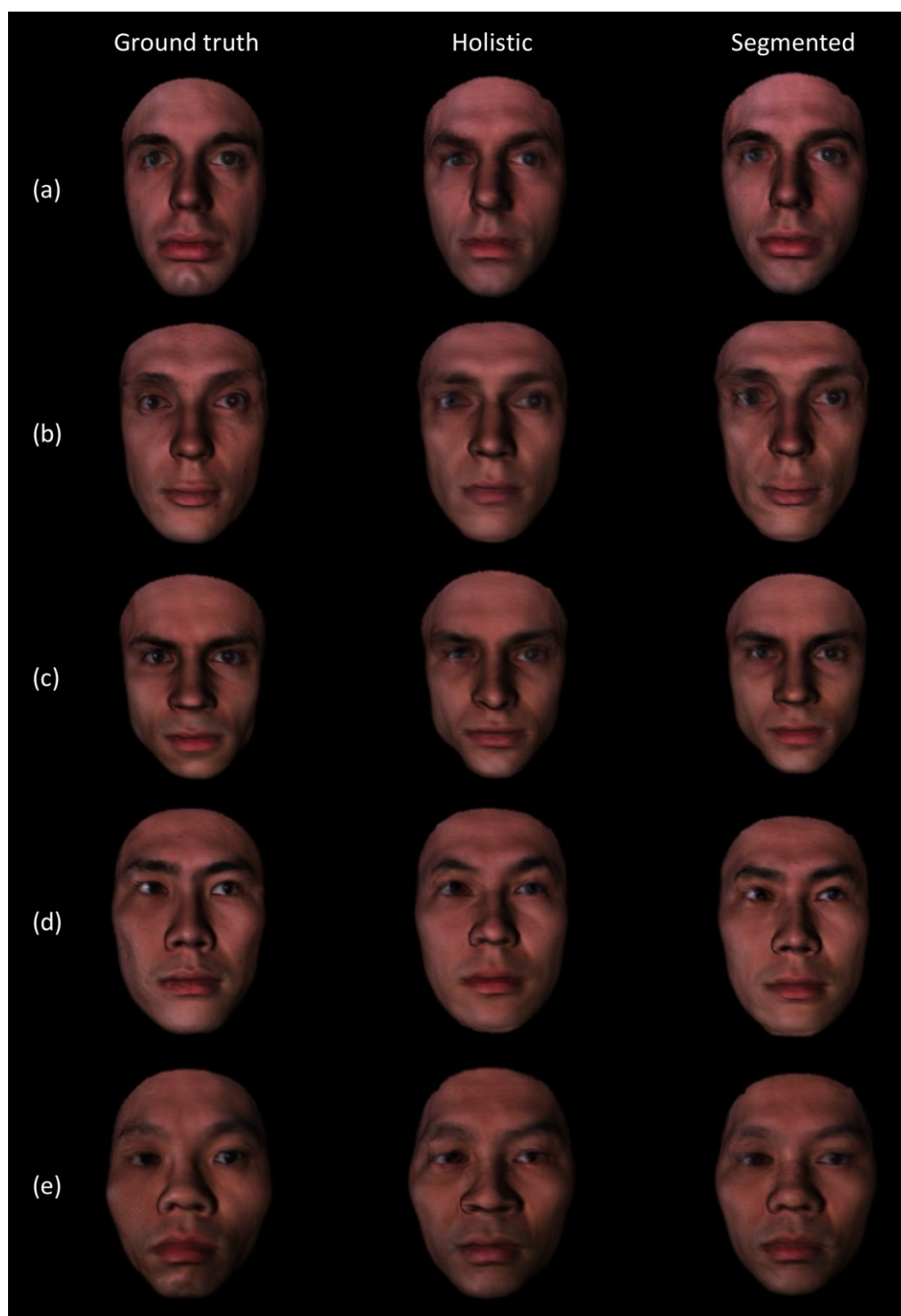
Figure 28. Face reconstruction. The figure shows several ground truth 3D models (a-e) and the corresponding reconstructions. Holistic reconstructions are shown in the middle column and the segmented-based reconstructions are illustrated in the last column.

Figure 26 shows a sample 3D model from the 3DFS dataset, and its corresponding shape and texture reconstruction based on PCA and our own implementation of Blanz and Vetter's iterative algorithm (see Appendix B and C). In turn, Figure 27 summarize reconstruction errors obtained through PCA back-projection against those obtained with Blanz and Vetter's iterative algorithm; as shown, PCA back-projection produced better reconstruction than the iterative algorithm. Finally, Figure 28 (middle column) shows examples of holistic-based reconstruction of several 3D face models from the 3DFS dataset.

6.2 Segment-wise reconstructions

Face reconstructions f'_{test} have 2×(m-1) degrees of freedom (m-1 associated with shape, and m-1 associated with texture). To increase the level of expressiveness, we segmented the face into four regions (Blanz and Vetter 1999), and performed the PCA decomposition in equations (28) and (29) for each segment independently; see Figure 29a. The final face model f'_{test} was obtained by combining each predicted segment using an image-blending algorithm (Burt and Adelson 1985) that works as follows. Given two input images (A and B) to be blended, we define a mask image (M) to denote whether the corresponding pixel should come from image A (mask value equal to 1) or B (zero). Then we construct a Laplace pyramid for images A and B, and a Gaussian pyramid for the mask image M, as illustrated in Figure 30. At each level in the
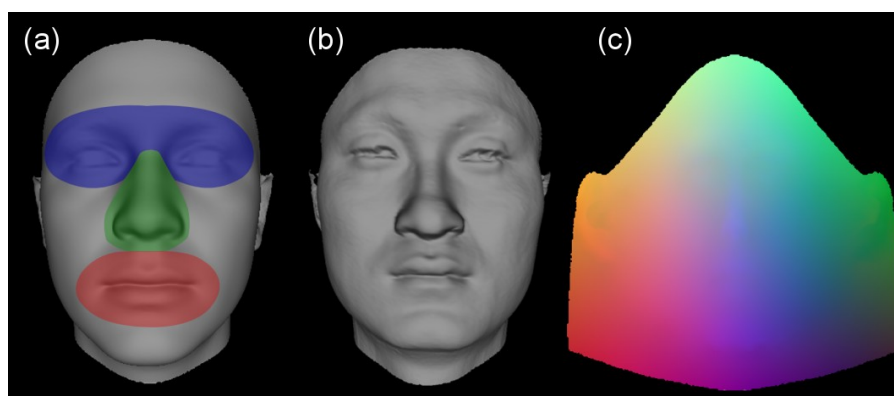
Figure 29. Face segments and geometry image samples. (a) Face segmentation used for 3D face reconstructions; each region is predicted independently and then merged into a composite face. (b) Example of a 3D shape and (c) its corresponding geometry image; the geometry image is an *n* × *m* matrix where XYZ coordinates are represented as RGB values.

$$G_n = \text{Reduce}(G_{n-1}) \qquad\qquad L_n = G_n - \text{Expand}(G_{n+1})$$
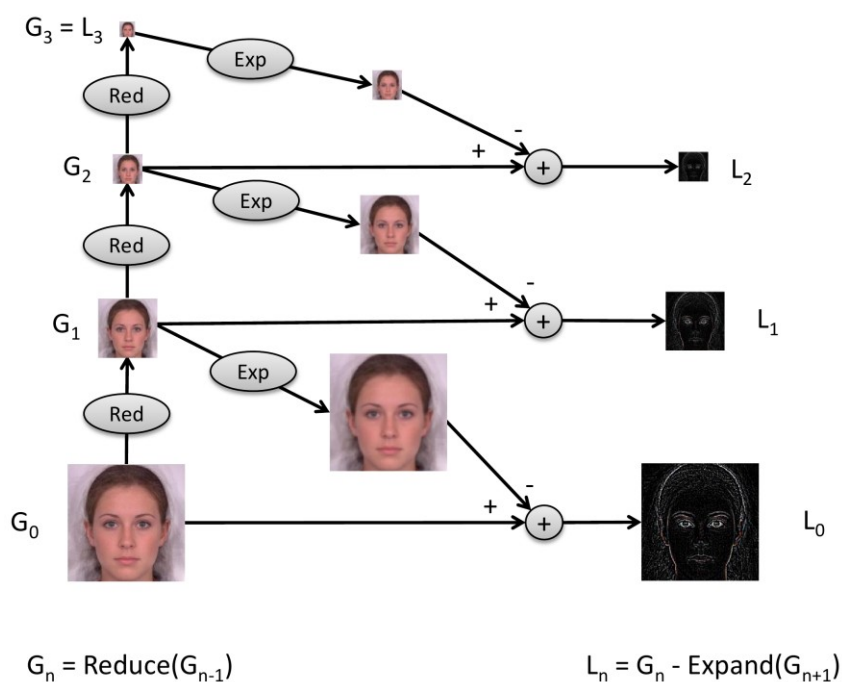
Figure 30. Construction of the Laplace pyramid. The process starts with a Reduce() operation (low-pass filtering) and sub-sampling (half sized) the original image, G_0, to obtain G_1. This process continues until reaching a predefined level N. G_N is known as the Gaussian pyramid. To complete the Laplace pyramid L_N, a band-pass operation is required between two successive low-pass levels. The lower frequency image (G_3 in the example), is interpolated using the Expand() operation before subtracting it from the higher frequency image, G_2. This process continues until reaching L_0. This figure was adapted from Burt and Adelson (1985).

pyramid, the algorithm blends the two images as:

$$LC_n(i,j) = GM_n(i,j) \cdot LA_n(i,j) + \big(1 - GM_n(i,j)\big) \cdot LB_n(i,j) \tag{34}$$

where LA_n, LB_n, and LC_n are the Laplace pyramids of the input images (A and B,) and the output image C, respectively, and GM_n is the Gaussian pyramid of the mask image for a given level n. Finally, the resulting blended image C (i.e. G_0), is synthesized from the LC_n pyramid as:

$$G_n = LC_n \tag{35}$$

$$G_{n-1} = LC_{n-1} + \text{Expand}(G_n) \tag{36}$$

In order to apply this image-based blending algorithm to 3D models, the 3D segments are converted into geometry images (GI) (Gu et al 2002) prior to the blending stage; see Figure 29b,c for an example of a geometry image from a 3D shape model. After blending all GI-based segments, the resulting GI is converted back into a 3D model. Reconstruction results with the segment-based method (f'_{test}) are illustrated in Figure 28 (last column). In turn, Figure 31 summarizes reconstruction errors with the segment-based and holistic-based reconstructions, which clearly show that segment-based reconstructions are significantly more accurate.
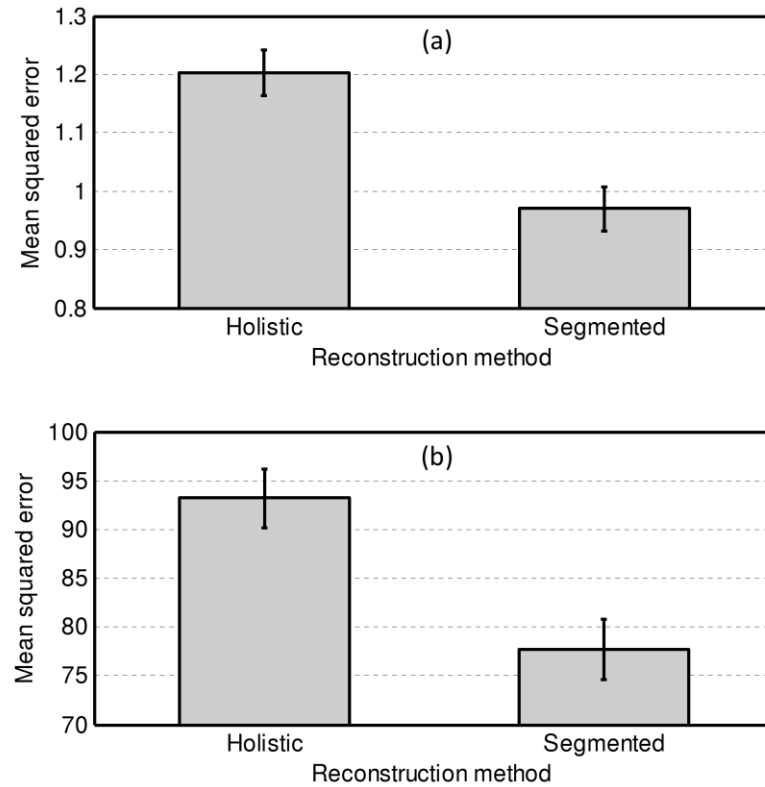
Figure 31. Mean squared error (MSE) per reconstruction method on the entire 3DFS dataset. Reconstruction was per formed in a leave-one-out (LOO) fashion. (a) MSE on shape reconstruction. (b) MSE on texture reconstruction. In both cases, the segmented-based method produced better reconstructions than the holistic-based method. Error bar are standard errors.

6.3 Caricaturization of segment-based reconstruction

Following the same caricaturization process described in Section 3, we generated

caricaturized 3D models from reconstructed test faces. Figure 32 illustrates several

samples of them at different caricaturization levels. The caricaturization process

enhanced the distinctive features of the face, but it seems that reconstruction artifacts
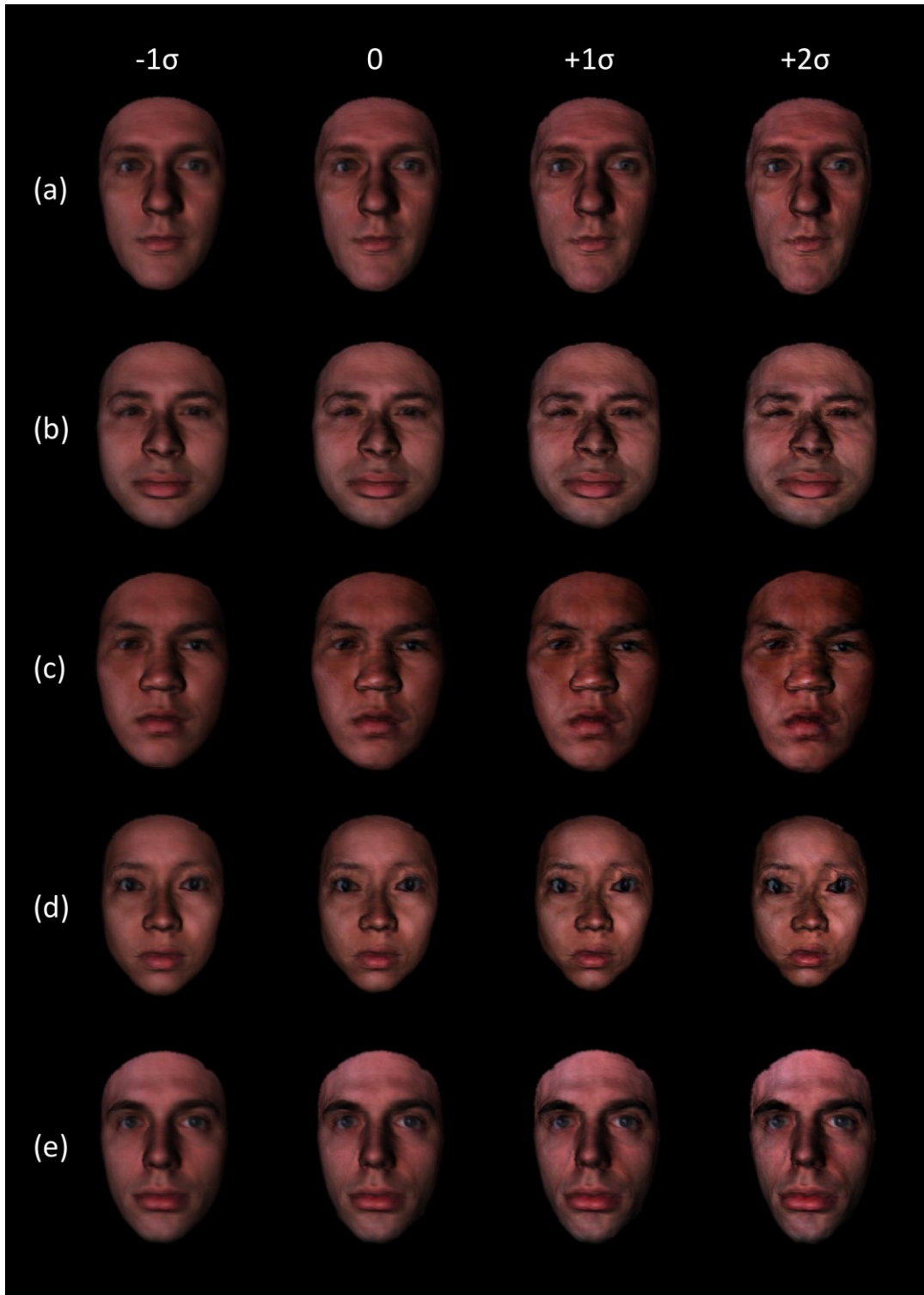
Figure 32. Caricatures from segmented reconstructions. This figure shows five different 3D reconstructed models at different caricaturization level from -1σ to +2σ.

are also amplified. This is clearly noticeable at higher caricaturization levels, see Figure 32's +2σ column. Whether these artifacts have an effect in face perception is the subject of the final perceptual experiment, which is described in Section 7.

7. PERCEPTUAL STUDY: RECOGNITION OF 3D FACE RECONSTRUCTIONS

7.1 Face perception using 3D model reconstructions

In this final perceptual study, we investigated whether caricatures generated from 3D reconstructions, as described in Section 6, would be valid stimuli for reverse caricature training. For this purpose, we employed the same old/new face recognition protocol described earlier in Sections 4 and 5, whereby subjects are first familiarized with a set of faces, and then asked to recognize those faces among a set of confounders. Following (Rodríguez et al 2008), we used the minimal degree of exaggeration between veridical and caricature faces that would lead to a caricature effect. This exaggeration level was then used in a recognition study that allowed us to test whether familiarization with caricatures of *reconstructed* 3D models (as opposed to caricatures of the *original* 3D models) would reduce OREs.

7.1.1 Stimuli

Forty face models were selected from the 3DFS-100 dataset (Blanz and Vetter 1999). The same 40 faces were used throughout the experiments. From these ground-truth veridical faces (V) (Figure 33a) we generated 40 veridical reconstructions (Vr) (Figure 33b), which were in turn used to generate 40 caricaturized reconstructions (Cr) (Figure 33d). Following Furl et al (Furl et al 2002), our face corpus had a similar distribution

Figure 33. Reconstruction studies sample stimuli. (a) ground-truth frontal faces, (b) their corresponding 3D segment-based reconstructions, (c) caricatures from ground-truth faces, and (d) caricatures from reconstructed faces. Ears and neck were manually removed to prevent participants from using picture-matching strategies. Inspection of (c) and (d) illustrates the extent to which caricatures amplify reconstruction errors rather than unique facial traits.

across races: 10 Caucasians, 10 East Asians, and 10 Indians. We also included 6 African faces and 4 faces from other groups (Middle Eastern, Hispanic) as filler stimuli. East Asian and Indian faces were treated as other-race faces.

7.1.2 Procedure

Forty-three Caucasian undergraduate students (24 females and 19 males) from the Department of Psychology at Texas A&M University participated in this study. Participants were assigned to one of two experimental conditions:

- Vr-V condition: familiarization with veridical reconstructions (Vr), recognition of veridical faces (V). Twenty-one students participated in this study, which served as a control.

- Cr-V condition: familiarization with caricaturized reconstructions (Cr), recognition of veridical faces (V). Twenty-two students participated in this condition, which tested our working hypothesis.

For each condition, participants were familiarized with 20 frontal target faces (the same faces for all subjects), each presented twice in random order, 3 seconds per presentation. Following familiarization, participants were tested on 40 faces, of which 20 were "new" (non-target) and 20 were "old" (target); all faces in the test phase were rendered with a random orientation between ±5 degrees in the three axes in order to prevent picture-matching strategies (Rodríguez et al 2009). Participants were asked to identify each face as "old" if they recognized it as one from the familiarization phase, or

as "new" otherwise. Following (Duchaine et al 2003), no time limits were imposed, but participants were asked to make the identification as rapidly as possible without sacrificing accuracy. In all conditions, the same 20 randomly chosen faces were used (in either caricature or veridical form). Following (Rodríguez et al 2008) exaggeration levels were set to -1σ ($\alpha = -0.21$) and 0σ ($\alpha = 0$) for veridical and caricaturized faces, respectively. A similar gender distribution for participants was maintained across experimental conditions.

7.1.3 Results

7.1.3.1 Do 3D reconstructions errors affect recognition performance?

To answer this question, we compared results from the Vr-V condition (i.e., in the new experiments above) against those on the V-V condition (i.e., from our earlier study in section 5.2 [15]) in terms of the signal detection d' measure (Macmillan and Creelman 2005). Results are summarized in Figure 34. Using the experimental condition (V-V vs. Vr-V) and race (Caucasian vs. East Asian vs. Indian) as factors, a two-way ANOVA shows a main effect of experimental condition, $F(2,82) = 8.735$, $p < 0.01$. Overall performance in the Vr-V condition is lower than in the V-V condition: $t(42) = 2.956$, $p < 0.01$, and also on each individual race. Thus, these results indicate that reconstruction errors have a negative effect on recognition performance, regardless of race. Looking at

_____

[15] Our earlier experiments used the same procedure, stimulus set, and subject pool as those reported here.
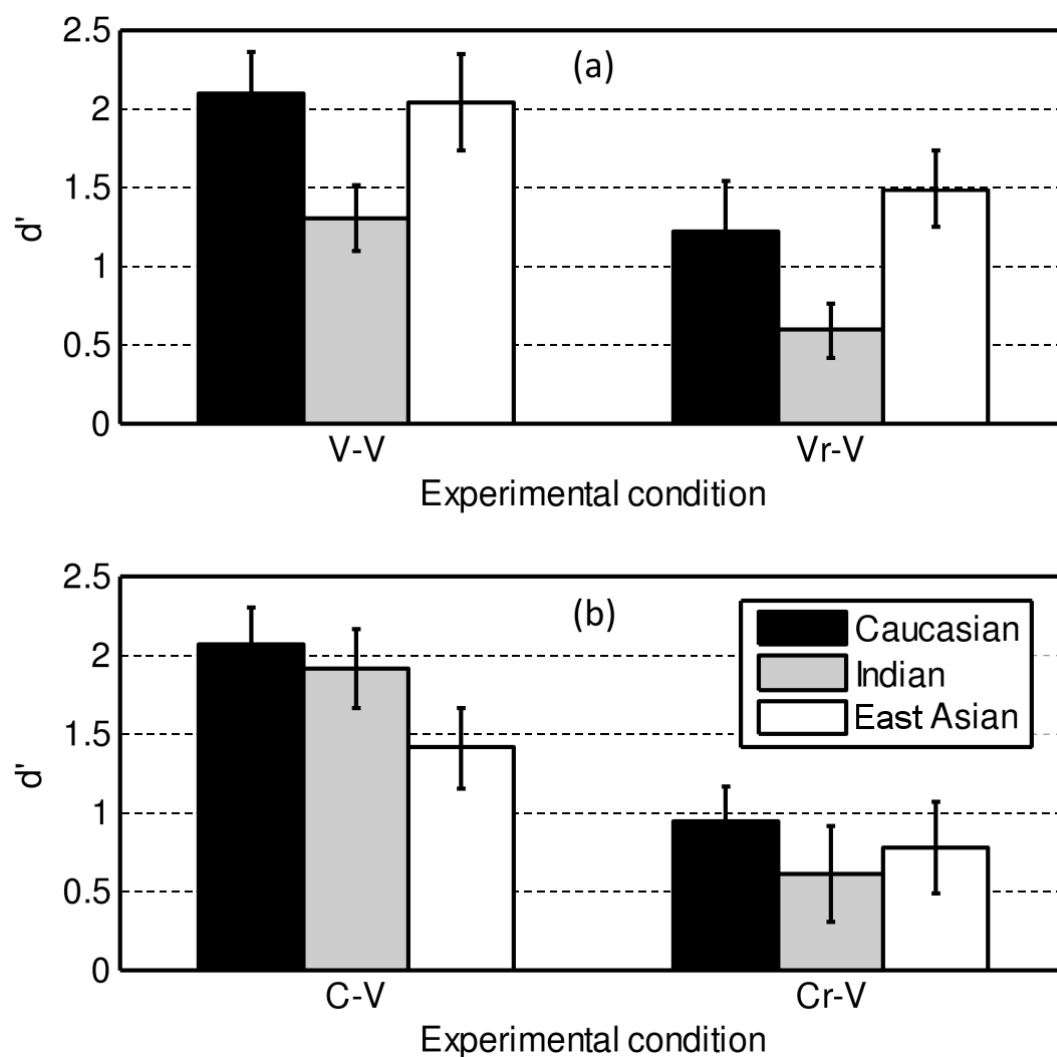
Figure 34. Signal detection d′ for (a) V-V and Vr-V conditions, and (b) C-V and Cr-V conditions. V-V and C-V conditions are based on Rodríguez et al (2008) results. Error bars represent standard errors.

participants' gender, experimental condition, and stimulus race as factor, there was not main effect of gender, $F(1,40) = 1.10$, *ns*. Namely, females and males recognition performance was similar.
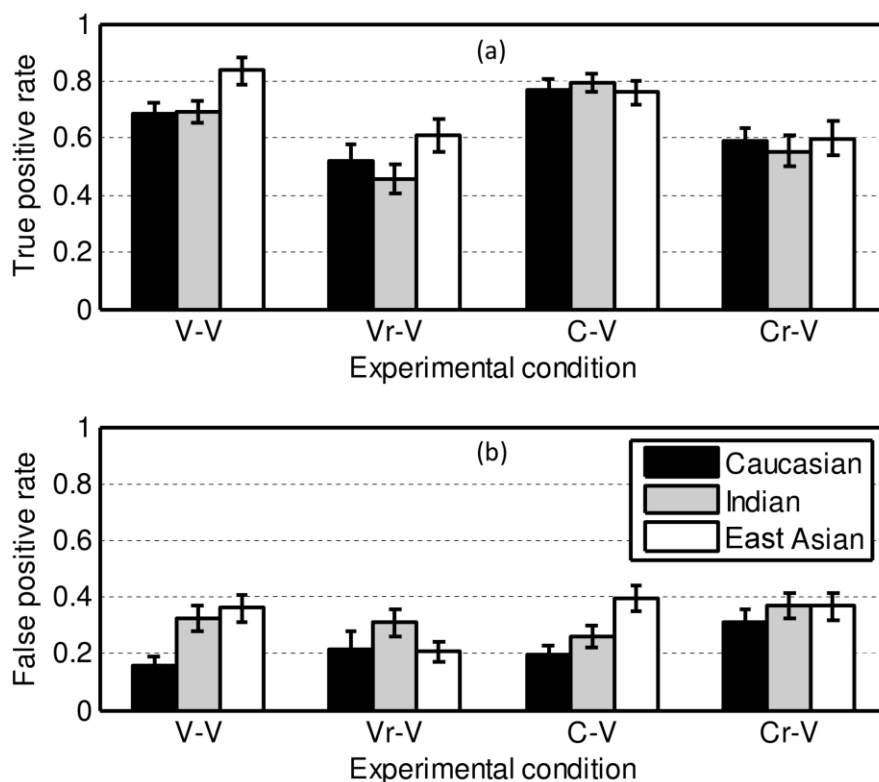
Figure 35. (a) True positive and (b) false positive rates across experimental conditions. V-V and C-V conditions are based on Rodríguez et al (2008) results. Error bars represent standard errors.

To tease apart the influence of reconstruction errors on different decisions, we also analyzed results in terms of true positive rates (TPR) and false positive rates (FPR). Results are summarized in Figure 35:

- TPRs showed a main effect of race $F(2,84) = 6.95$, $p < 0.01$, and of experimental condition, $F(1,42) = 18.108$, $p < 0.001$. There was no interaction effect between race and experimental condition, $F(2,84) = 0.444$, *ns*. Overall performance on the V-V condition was significant higher than on the Vr-V condition, $t(42) = 4.255$, $p < 0.001$.

- FPRs showed a main effect of race $F(2,84) = 4.99$, $p < 0.05$, and interaction effects between race and experimental condition, $F(2,84) = 3.125$, $p < 0.05$. There is no main effect of experimental condition, $F(1,42) = 0.687$, $p > 0.05$. Overall performance on the V-V condition was not significant different than on the Vr-V condition, $t(42) = 0.829$, $p > 0.05$.

These results indicate that reconstruction errors affected recognition performance by decreasing TPRs but not necessarily by increasing FPRs. However, there was a main effect of race on both measures, which suggests that reconstruction errors affect the viewer's performance differently depending on the race of the stimulus face.

Finally, we analyzed the signal detection criterion C, which provides cues about whether participants have a bias toward a particular answer. A conservative participant will answer 'no' more often (positive C values), while a liberal participant will respond 'yes' more often (negative C values) (Abdi 2007). Results are shown in Figure 36a. Analysis of variance shows a main effect of race: $F(2,84) = 7.55$, $p < 0.001$ and a main effect of experimental condition: $F(1,42) = 9.621$, $p < 0.01$, but no interaction effects: $F(2,84) = 2.914$, ns. Participants on the V-V condition were more conservative to own-race faces than to other-race faces. In addition, training on the Vr-V condition caused a noted increase in criterion C, which indicates that reconstruction errors made participants more conservative. We can infer participants were having difficulties learning from reconstructions. This is consistent with the TPRs results above;
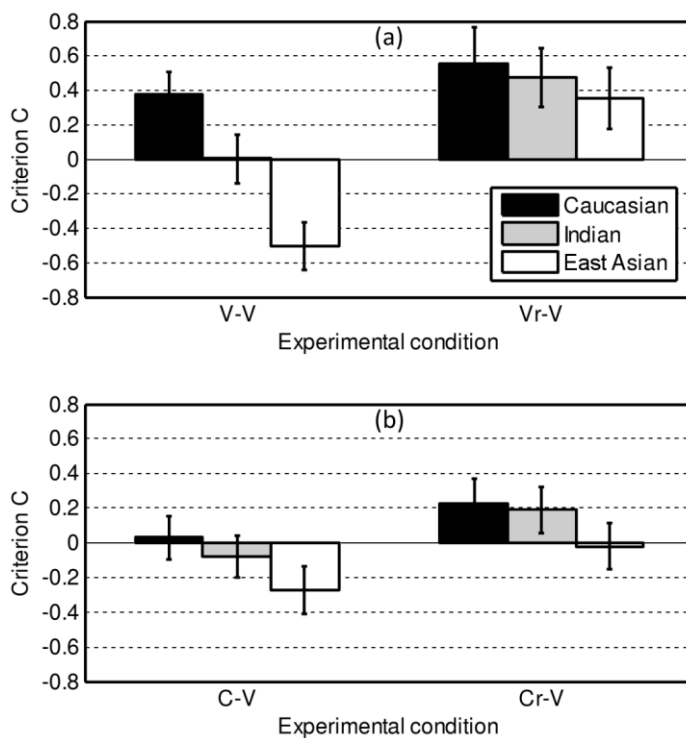
Figure 36. Criterion C across experimental conditions. V-V and C-V conditions are based on Rodríguez et al (2008) results. Error bars represent standard errors.

conservative participants would experience a reduced TPR score. Furthermore, conservative participants would improve their FPRs. In this pair of experimental conditions, participants maintained the same level of FPR scores.

7.1.3.2 Do caricatures reduce the impact of 3D reconstruction errors on recognition performance?

To answer this question, we compared results in the Vr-V and Cr-V conditions, also in terms of the signal detection d' measure. Using experimental condition (Vr-V vs. Cr-V) and race (Caucasian vs. East Asian vs. Indian) as factors, a two-way ANOVA shows no

main effect of experimental condition, $F(1,41) = 2.068$, $p > 0.05$. However, performance

was lower on Cr-V than on Vr-V: $t(41) = 1.438$, *ns*, ($p = 0.15$). There was no main effect

of race, $F(2,82) = 2.519$, *ns*. OREs on Indian faces were observed for participants in the

Vr-V condition: $t(40) = 1.69$, *ns* ($p = 0.099$), but not for participants in the Cr-V condition:

$t(42) = 0.88$, *ns*. However, this reduction of OREs with reverse caricatures was due to a

reduction in recognition performance on Caucasian faces rather than to an increase in

recognition performance on Indian faces. There was no significant interaction between

race and training condition, $F(2,82) = 0.997$, *ns*; both training conditions show similar

performance profile across races. Therefore, when considering the signal detection d′

measure, reverse-caricature training using reconstructed 3D models did not reduce the

impact of reconstruction errors. In fact, it appears that the caricature process amplifies

reconstruction errors to a greater extent that it makes the unique facial traits more

salient. Looking at participants' gender, experimental condition, and stimulus race as

factor, there was not main effect of gender, $F(1,39) = 1.25$, *ns*. Namely, females and

males recognition performance was similar.

As before, we performed a finer-grained analysis in terms of true positive rates

(TPR) and false positive rates (FPR). Results are summarized in Figure 35a-b:

- TPRs did not show a main effect of race $F(2,82) = 2.096$, *ns*, training condition, $F(1,41) = 0.959$, *ns*, or interaction effects, $F(2,82) = 0.681$, *ns*. However, performance was better on Cr-V than it was on Vr-V: $t(41) = 0.972$, $p > 0.05$.

- FPRs did not have a main effect of race: $F(2,82) = 1.275$, *ns*, but had a significant main effect of condition: $F(1,41) = 6.332$, $p < 0.05$. Namely, performance was worse on Cr-V than on Vr-V: $t(41) = 2.516$, $p < 0.05$.

These results suggest that although the caricature process may improve TPRs, this is at the expense of a much larger increase in FPRs such that the net effect of caricaturization (as measured by the signal detection d' measure) is detrimental.

Finally, we also analyzed the signal detection criterion C. Results are shown in Figure 36. Analysis of variance shows no main effect of race, $F(2,82) = 1.424$, *ns*, and no significant interaction, $F(2,82) = 0.050$, *ns*. However, there is a main effect of training condition $F(1,41) = 3.80$, $p < 0.05$. Namely, participants in the Cr-V condition become less conservative than those in the Vr-V condition and approach the ideal observer (i.e., C=0, no strategy).

7.1.3.3 Do caricatures of reconstructed faces provide better recognition performance than veridical faces?

To answer this question, we compared results in the Cr-V and V-V conditions in terms of the signal detection d' measure. Using experimental condition (Cr-V vs. V-V) and race (Caucasian vs. East Asian vs. Indian) as factors, a two-way ANOVA did not show a main effect of race, $F(2,86) = 2.796$, *ns*, and did not have an interaction between race and experimental condition. There was a main effect of experimental condition, $F(1,43) = 17.792$, $p < 0.001$; performance on the V-V condition was significantly higher than on the

Cr-V condition, $t(43) = 4.218$, $p < 0.001$. Thus, the reverse-caricature effect we observed in our earlier work (Rodríguez et al 2008) disappears when caricatures are obtained from reconstructed 3D models, which suggests that reconstruction errors can perceptually mask distinctive facial cues. Looking at participants' gender, experimental condition, and stimulus race as factor, there was not main effect of gender, $F(1,41) = 0.18$, *ns*. Namely, females and males recognition performance was similar.

As before, we also analyzed TPRs and FPRs; results are summarized in Figure 35:

- TPRs did not show a main effect of race $F(2,86) = 2.512$, *ns*, and no interaction effects, $F(2,86) = 1.313$, *ns.* TPRs showed a main effect of training condition, $F(1,43) = 12.09$, $p < 0.001$. Namely, Cr-V performance was significantly lower than V-V performance: $t(43) = 3.477$, $p < 0.001$.

- FPRs had a main effect of race, $F(2,86) = 5.501$, $p < 0.01$, but no effect of condition, $F(1,43) = 3.184$, *ns*, or interaction effects: $F(2,86) = 1.759$, *ns*.

These results mirror those in subsection 7.1.3.1 and indicate that the lower recognition performance on Cr-V is due to a reduction of TPRs rather than an increase in FPRs.

Finally, analysis of variance on the signal detection criterion C, summarized in Figure 36, shows no main effect of experimental condition: $F(1,43) = 2.204$, *ns*, or interaction effects: $F(2,86) = 2.799$, *ns*. However, there is a main effect of race: $F(2,86) = 9.031$, $p < 0.001$, namely in terms of reduced OREs in the Cr-V condition.

7.2 Summary

In Sections 4 and 5, we presented perceptual studies that caricatures from ground-truth 3D models improve the recognition of their veridical counterparts and also reduce OREs, as measured by the signal detection's sensitivity index d' (Macmillan and Creelman 2005). The main objective of this final study was to determine whether these earlier results would extend to face recognition when the veridical 3D models are not available and have to be replaced by reconstructions.

Our results indicate that 3D reconstructions are not of sufficient quality to be used for face recognition purposes, even when rendered without caricaturization. Training with reconstructed faces (Vr-V) leads to lower TPRs when compared to training with veridical faces (V-V), although FPRs seem immune to reconstruction errors. Training on caricatures of reconstructed faces (Cr-V) leads to higher TPRs when compared to training on reconstructed faces (Vr-V) but at the expense of a larger increase in FPRs, with a negative net effect. This suggests that the caricature process amplifies reconstruction errors more than it enhances distinctive facial features. Finally, training with caricatures of reconstructions (CrV) leads to lower recognition performance than training on veridical faces (V-V), mainly in terms of reduced TPRs.

Collectively, our study indicates that the reconstruction process fails to capture the more distinctive features of a given face (e.g., notice the missing chin dimple in the second row of Figure 33). Because caricatures amplify differences relative to a norm, they also exacerbate any errors introduced during reconstruction (e.g., first row face in

Figure 33), possibly distracting participants away from the distinctive features of faces. Thus, our results suggest that the original 3D faces may be required in order to generate perceptually valid caricatures, and that one needs to pay close attention to the reconstruction process, specifically, the quality of synthesized face and its impact on face perception (Rodríguez and Gutierrez-Osuna 2011). This reconstruction process could introduce unwanted features to the extent that it may prevent individuals from focusing their attention on features that are useful for recognition. Additionally, the reconstructed faces might not represent accurately the distinctive features, affecting the recognition performance and reducing the effectiveness of caricatures as a training aid.

8. SUMMARY

By definition, caricatures increase the salience of idiosyncratic or normatively distinct qualities. This form of distortion appears to increase the amount of memorable information available for later recognition. As a result, exposure to facial caricatures can increase later recognition of their veridical counterparts. Furthermore, these distortions appear to decrease the own-race recognition advantage, possibly by allowing participants to focus on those features that are more distinctive for other-race faces. This apparent paradox has important implications for the development of training tools for face recognition.

Although a caricature advantage has been demonstrated in previous work, many of these investigations focused on cases in which faces were exaggerated during both familiarization and recognition, e.g. (Benson and Perrett 1994; Rhodes et al 1987). Effects of this sort are theoretically important, but it is unclear how they might be applied to enhance memory of faces in practice, when most operational scenarios demand that faces be recognized in their veridical form. Demonstrations of a reverse-caricature effect similar to that found in Section 4 (e.g., caricaturized training; veridical testing) were based on very different stimuli, such as line drawings and veridical photographs (Stevenage 1995), or three-quarter view representations of faces (Deffenbacher et al 2000).

In the initial perceptual study, described in Section 4, we tested whether the "reverse-caricature effect" would be robust to procedural variations that created more difficult learning environments. Specifically, we examined whether the effect would emerge with frontal rather than three-quarter views, after very brief exposure to caricatures during the learning phase and after modest rotations of faces during the recognition phase. Participants were assigned to one of three experimental conditions: (*i*) veridical familiarization/veridical test (VV), (*ii*) caricature familiarization/caricature test (CC), and (*iii*) caricature familiarization/veridical test (CV). VV and CC were our control groups and set up the lower and upper boundary for CV, our experimental group. Participants on the CC condition showed higher performance than the participants on the VV condition, a result that is consistent with the perceptual caricature effect reported in the literature. Participants on the CV condition showed a significant improvement over the VV results, but lower than the results obtained on the CC experimental condition. Results indicate that, even under these difficult training conditions, people are more accurate at recognizing unaltered faces if they are first familiarized with caricatures of the faces, rather than with the unaltered faces (Rodríguez et al 2009). Our findings extend previous reports of a reverse-caricature effect, further demonstrating its potential as a training tool for face recognition. In contrast to the work by Deffenbacher et al (2000), in which an optimal three-quarter view and a 60-second familiarization exposure were used, our study shows that a reverse-caricature effect can also be obtained when only a frontal representation is

available during training, with a significantly shorter familiarization time, and that the effect is robust to modest rotations of faces. These results support the use of caricatures as part of a training tool for face recognition in applied settings. Consider law enforcement as one promising area. Police officers often are required to learn the faces of the at-large suspects that they must try to apprehend. Law enforcement agencies also post the pictures of wanted individuals when they are soliciting the assistance of the public.

How could the results from this study be extended to these applied settings? First, stimuli would have to contain both the 3-D shape and the 2-D reflectance/texture of faces. Our focus on facial shape in the first study (Section 4) served two purposes: it helped participants focus on anthropometric features of the face rather than on non-configural features (e.g., skin color, skin blemishes, eye color), and made our results comparable with those of Deffenbacher et al (2000), who also used shape-only faces. Extensions of the caricature process to both facial shape and facial reflectance are technically straightforward; the challenge lies in determining suitable levels of exaggeration for shape and reflectance when both sources of information are present. To this end, work by O'Toole et al (1999) suggests that the relative contribution of shape and texture may be gender dependent; recognition of male faces depends more on shape than on reflectance, whereas recognition of female faces relies on both channels equally. Furthermore, reflectance (i.e., skin color in this context) is one of the important traits in race determination (Jingrong et al 2004), bringing the other-race perceptual

effect to fore. We included reflectance in the study in Section 5 and investigated the presence of other-race effects and to determine the extent to which familiarization with caricaturized faces could also be used to reduce other-race effects. For this purpose, Caucasian participants were first familiarized with a set of faces from multiple races, and then asked to recognize those faces among a set of confounders. Specifically, participants were assigned to one of two experimental conditions: (*i*) veridical familiarization-veridical test (VV), which served as a control group, and (*ii*) caricature familiarization-veridical test (CV), which tested our hypothesis that reverse-caricature training can reduce other-race effects. Participants on VV showed a significant other-race effect on Indian faces, whereas participants on CV showed no other-race effects on Indian faces. This result suggests that caricaturization may be used to help individuals focus their attention to features that are useful for recognition of other-race faces (Rodríguez et al 2008).

Second, the caricature process requires a 3-D model for each target face. In most practical settings, however, one will not have the luxury of performing a 3-D face scan of each target face; 3-D scanners are still expensive instruments, and scanning is not possible when the wanted individual is at large. One potential solution to this problem is to use photogrammetric techniques to reconstruct 3-D face models from 2-D photographs; visually convincing reconstructions have been demonstrated by Blanz and Vetter (1999) using morphable models. Whether these reconstructions are sufficiently accurate to be used for caricaturing and training is a matter that required

further investigation, as even minute reconstruction errors would be amplified by the exaggeration process. We studied this particular issue in Sections 6 and 7, where we sought to determine the extent to which familiarization with 3D reconstructed caricaturized faces could also be used to reduce other-race effects. Using PCA and an image blending algorithm, we obtained a 3D approximation of a probe face from a gallery of 3D faces. These reconstructions were then caricaturized. Caucasian participants were first familiarized with a set of faces from multiple races, and then asked to recognize those faces among a set of confounders. Specifically, participants were assigned to one of two experimental conditions: (*i*) familiarization with veridical reconstructions - veridical test (VrV), (*ii*) familiarization with caricaturized reconstructions - veridical test (CrV). Participants who were familiarized with reconstructed faces and then asked to recognize the ground truth versions of the faces showed a significant reduction in performance compared to the VV results described in Section 5. In addition, participants who were familiarized with caricatures of reconstructed versions, and then asked to recognize their corresponding ground truth versions, showed a larger reduction in performance. These results are critical for the development of training tools because in most realistic settings 3D scans of the target faces are not available. In these cases, all that is available is a 2D "mug shot" of the target. While at the onset of this study we anticipated that caricatures would amplify reconstruction errors, it was not clear whether these errors would compromise human's ability to recognize faces. As an example, humans recognize faces under fairly severe

manipulations (e.g., at very low resolution, under partial occlusion), but this ability breaks down with other types of lossless manipulations (e.g., rotated or inverted faces); see (Sinha et al 2006).

8.1 Conclusions

The main objective of this dissertation was to determine the extent to which photorealistic computer-generated caricatures could be used in training tools to improve recognition of faces by individuals. This research offer clear evidence that facial distortions can improve face encoding and subsequent recognition, if they exaggerate features that make a face most distinctive with regard to a particular population of faces. We showed that caricatures are useful to improve overall recognition, including reducing other-race effects. Caricatures simply allowed us to focus our attention to critical features for recognition. These results are encouraging for the development for face recognition training tools for law enforcement or for medical fields (e.g., learning tools for people with autism, or people with prosopagnosia.) However, we also found that errors introduced during 3D face reconstructions significantly affected recognition, and that caricatures of reconstructed faces further exacerbated recognition errors. Reducing the impact of these reconstruction artifacts is an important area for further research.

8.2 Future directions

Future research will need to explore the cognitive mechanisms underlying the reverse-caricature effect and identify the boundary conditions that guide its application. The effects of race, gender and age on reverse-caricature training would have to be characterized in order to design large facial databases. In Section 4 study, we presented familiarization and test faces one at a time on a computer and, as a result, participants were able to gain some sense of the norms that defined the full population. This might be necessary for the type of effects we observed, as caricatures can only be defined in relation to the norms of a given population. Perhaps if we had used other methods in which the stimuli were presented in isolation or embedded within a sample of faces that had different norms (e.g., other-race effects), our caricature effect would have been diminished or disappeared altogether. These ambiguities point to the need for future research that can define the boundary conditions for our effect, so that caricatures can be used successfully to improve face recognition in real world settings.

Based on results from the third perceptual experiment, described in Section 7, we may be tempted to infer that better reconstruction algorithms would be needed, such as the non-linear optimization method in Blanz and Vetter (1999). While this may be the case in more general reconstruction scenarios (e.g., with real photographs), the reconstruction method in equations (32) and (33) is optimal (in the mean-square-error sense) for our study since the representation is linear in the optimization parameters (shape and texture) and all remaining parameters (camera and illumination) were

assumed known. Thus, failure to reproduce results in Sections 4 and 5 (which used ground-truth 3D faces) must be attributed to factors other than the reconstruction procedure, and instead point in the direction of facial databases. Specifically, extending the 3D face database would certainly improve the reconstruction quality, as additional faces would provide a more diverse pool from which to reconstruct new faces. Likewise, increasing the number of input images of a probe face (e.g. frontal, 3/4 and profile views) might help capture distinctive facial traits that are not prominent in a frontal view. Facial diversity (race, gender, age) most likely plays a significant role in the quality of reconstruction. Having a well-balanced database or one that specifically matches the characteristics of the input face (e.g. Caucasian, male, 55-60 years of age) might also improve reconstruction results. Additional research is also needed to determine, among others, (*i*) the level of reconstruction accuracy that must be achieved to obtain *perceptually* realistic (as opposed to photorealistic) results, (*ii*) the type and number of facial segments (Figure 29a), maybe on a race-by-race basis or to account for facial asymmetries, and (*iii*) the types of reconstruction errors (e.g. shape vs. reflectance, different facial areas) that have the greatest impact on recognition performance. Addressing these questions is a necessary step towards the development of effective training tools for face recognition.

We also evaluated whether reverse-caricatures can improve the recognition of *specific* other-race faces, i.e. a closed-world assumption. While we believe that, with sufficient exposure to caricaturized other-race faces, reduction of OREs may generalize

to faces not seen in the familiarization phase (see Hills and Lewis 2006), this issue will also require further investigation. However, in security applications where one seeks to improve the recognition of specific faces (i.e. those of a suspect at large), the closed-world assumption is valid.

Reverse-caricature effects can also be explored in the context of automatic face recognition. Can similar training strategies be applied to existing face recognition algorithms? For instance, caricatures may assist existing algorithms on local feature selection or model initialization by highlighting distinctive facial regions.

Finally, an important extension of this research is the development of computer models that accounts for all the perceptual effects discussed in this dissertation. We can find several models in the literature (Caldara and Abdi 2006; Haque and Cottrell 2005; Lewis 2004; O'Toole et al 1991; O'Toole et al 2001), but they address specific perceptual effects. Lewis (2004) accounts for distinctiveness, caricature, and race effects, but the model fails on reverse-caricature scenarios as we highlighted in subsection 2.2.2. O'Toole et al (1991), Caldara and Abdi (2006), and Haque and Cottrell (2005) dealt specifically with other-race effects. However, unified computational models that account for caricature effects, other-race effects, reverse-caricature effects and reversibility of the other-effects are yet to be developed. A unified model would help us understand how we process and learn faces, and more importantly how can we devise strategies to improve our recognition performance. Finally, this unified model

would allow us to make predictions about boundary conditions where perceptual

effects would be valid.

REFERENCES

Abdi H, 2007 "Signal detection theory", in *Encyclopedia of Measurements and Statistics* Ed N Salkind (Thousand Oaks, CA: Sage) pp 886-889

Amin S H, Gillies D, 2007 "Analysis of 3D face reconstruction", in *14th International Conference on Image Analysis and Processing* (Modena, Italy: IEEE Computer Society) pp 413-418

Atick J J, Griffin P A, Redlich A N, 1997 "Statistical approach to shape from shading: reconstruction of 3D face surfaces from single 2D images" *Neural Computation* **8** 1321-1340

Atkinson R C, Shiffrin R M, 1968 "Human memory: A proposed system and its control processes", in *The Psychology of Learning and Motivation* Eds K W Spence and J T Spence (New York: Academic Press) pp 89–195

Benson P J, Perrett D I, 1991a "Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images" *European Journal of Cognitive Psychology* **3** 105-135

Benson P J, Perrett D I, 1991b "Synthesizing continuous-tone caricatures" *Image and Vision Computing* **9** 123-129

Benson P J, Perrett D I, 1994 "Visual processing of facial distinctiveness" *Perception* **23** 75-93

Blanton H, Christie C, 2003 "Deviance regulation: A theory of identity and action" *Review of General Psychology* **7** 115-149

Blanz V, O'toole A J, Vetter T, Wild H A, 2000 "On the other side of the mean: The perception of dissimilarity in human faces" *Perception* **29** 885-891

Blanz V, Vetter T, 1999 "A morphable model for the synthesis of 3D faces", in *26th Annual Conference on Computer Graphics and Interactive Techniques* (Los Angeles, CA: ACM Press/Addison-Wesley) pp 187-194

Blanz V, Vetter T, 2002 "Face recognition based on 3D shape estimation from single images", in *Technical Report CGF-TR 2* (Freiburg, Germany: University of Freiburg)

Blanz V, Vetter T, 2003 "Face recognition based on fitting a 3D morphable model" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25** 1063-1074

Bothwell R K, Brigham J C, Malpass R S, 1989 "Cross-racial identification" *Personality and Social Psychology Bulletin* **15** 19-25

Brennan S E, 1985 "Caricature generator: The dynamic exaggeration of faces by computer" *Leonardo* **18** 170-178

Brigham J C, Barkowitz P, 1978 "Do they all look alike? The effect of race, sex, experience, and attitudes on the ability to recognize faces" *Journal of Applied Social Psychology* **8** 306-318

Brigham J C, Bennett L B, Meissner C A, Mitchell T L, 2007 "The influence of race on eyewitness memory", in *Handbook of Eyewitness Psychology: Memory for People* Eds

R Lindsay, D Ross, J Read and M Toglia (London, UK: Lawrence Erlbaum & Associates) pp 257-281

Brigham J C, Malpass R S, 1985 "The Role of experience and contact in the recognition of faces of own- and other-race persons" *Journal of Social Issues* **41** 139-155

Bruce V, Valentine T, Baddeley A D, 1987 "The basis of the 3/4 view advantage in face recognition" *Applied Cognitive Psychology* **1** 109-120

Burt P J, Adelson E H, 1985 "Merging images through pattern decomposition", in 8th *SPIE Proceedings on Applications of Digital Image Processing* Ed A G Tescher (Bellingham, WA: Society for Photo-Optical Instrumentation Engineers) pp 173-181

Caldara R, Abdi H, 2006 "Simulating the other-race effect with autoassociative neural networks: Further evidence in favor of the face-space model" *Perception* **35** 659-670

Chance J E, Goldstein A G, 1981 "Depth of processing in response to own- and other-race faces" *Personality and Social Psychology Bulletin* **7** 475-480

Chance J E, Goldstein A G, 1996 "The other-race effect and eyewitness identification", in *Psychological Issues in Eyewitness Identification* Eds S L Sporer, R S Malpass and G Koehnken (Mahawah, NJ: Lawrence Erlbaum Associates) pp 153-165

Chance J E, Goldstein A G, McBride L, 1975 "Differential experience and recognition memory for faces" *The Journal of Social Psychology* **97** 243-253

Cheng C M, Lai S H, 2001 "An integrated approach to 3D face model reconstruction from video", in *Proceedings of the 2001 IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems* (Vancouver, BC: IEEE) pp 16-22

Cheng Y D, Knappmeyer B, Bülthoff I, 2000 "The caricature effect across viewpoint changes in face perception", in *8th Annual Workshop on Object Perception and Memory* (New Orleans, LA: Psychonomic Society)

Choi K H, Hwang J N, 2002 "A Real-Time System for Automatic Creation of 3D Face Models from a Video Sequence", in *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Orlando, FL: IEEE Signal Processing Society) pp 2121-2124

Deffenbacher K A, Johanson J, Vetter T, O'Toole A J, 2000 "The face typicality-recognizability relationship: Encoding or retrieval locus?" *Memory & Cognition* **28** 1173-82

Duchaine B, Nieminen-von Wendt T, New J, Kulomaki T, 2003 "Dissociations of visual recognition in a developmental agnosic: Evidence for separate developmental processes" *Neurocase* **9** 380-389

Duda R O, Hart P E, Stork D G, 2001 *Pattern classification, Second Edition* (New York, NY: Wiley)

Ellis H, Deregowski J, Shepherd J, 1975 "Description of white and black faces by white and black subjects" *International Journal of Psychology* **10** 119-123

Fan X, Peng Q, Zhong M, 2010 "3D face reconstruction from single 2D image based on robust facial feature points extraction and generic wire frame model", in *Proceedings of the 2010 International Conference on Communications and Mobile Computing* (Shenzhen, China: IEEE Computer Society) pp 396-400

Fanany M I, Kumazawa I, 2002 "Analysis of shape from shading algorithms for fast and realistic 3D face recognition", in *Proceedings of the 2002 Asia-Pacific Conference on Circuits and Systems* (Singapore: IEEE) pp 181-185

Fanany M I, Ohno M, Kumazawa I, 2002 "A scheme for reconstructing face from shading using smooth projected polygon representation NN", in *Proceedings of the 2002 International Conference on Image Processing* (Rochester, NY: IEEE) pp 305-308

Forster K I, Forster J C, 2003 "DMDX: A Windows display program with millisecond accuracy" *Behavior Research Methods, Instruments, & Computers* **35** 116-124

Fukunaga K, 1990 *Introduction to Statistical Pattern Recognition, Second Edition* (San Diego, CA: Academic Press)

Furl N, Phillips P J, O'Toole A J, 2002 "Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis" *Cognitive Science* **26** 797-815

Going M, Read J D, 1974 "The effect of uniqueness, sex of subject, and sex of photograph on facial recognition" *Perceptual and Motor Skills* **39** 109-110

Goldstein A G, 1979a "Facial feature variations: Anthropometric data II" *Bulletin of the Psychonomic Society* **13** 191-193

Goldstein A G, 1979b "Race-related variation of facial features: Anthropometric data I" *Bulletin of the Psychonomic Society* **13** 187-190

Goldstein A G, Chance J E, 1976 "Measuring psychological similarity of faces" *Bulletin of the Psychonomic Society* **7** 407-408

Goldstein A G, Chance J E, 1980 "Memory for faces and schema theory" *The Journal of Psychology* **105** 47-59

Goodman G S, Sayfan L, Lee J S, Sandhei M, Walle-Olsen A, Magnussen S, Pezdek K, Arredondo P, 2007 "The development of memory for own- and other-race faces" *Journal of Experimental Child Psychology* **98** 233-242

Goodman N D, Tenenbaum J B, Feldman J, Griffiths T L, 2008 "A rational analysis of rule-based concept learning" *Cognitive Science* **31** 108-154

Gu X, Gortler S, Hoppe H, 2002 "Geometry images", in *29th Annual Conference on Computer Graphics and Interactive Techniques* (San Antonio, TX: ACM Press) pp 355-361

Guan Y, 2007 "Automatic 3D face reconstruction based on single 2D image", in *Proceedings of the 2007 International Conference on Multimedia and Ubiquitous Engineering* (Seoul, South Korea: Science & Engineering Research Support Center) pp 1216-1219

Haque A, Cottrell G W, 2005 "Modeling the other-race advantage with PCA", in *27th Annual Conference of Cognitive Science Society* (La Stresa, Italy: Cognitive Science Society) pp 899-904

Hassner T, Basri R, 2006 "Example based 3D reconstruction from single 2D images", in *2006 Computer Vision and Pattern Recognition Workshop* (New York, NY: IEEE Computer Society)

Hernandez B, 2006 "2D GI rendering tool (Version 1.0)", (Mexico City, Mexico: Instituto Tecnológico y de Estudios Superiores de Monterrey)

Hills P J, Lewis M B, 2006 "Reducing the own-race bias in face recognition by shifting attention" *The Quarterly Journal of Experimental Psychology* **59** 996-1002

Hopfield J, 1982 "Neural networks and physical systems with emergent collective computational abilities" *Proceedings of the National Academies of Science* **79** 2554–2558

Hu Y, Jiang D, Yan S, Zhang L, Zhang H, 2004 "Automatic 3D reconstruction for face recognition", in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition* (Southampton, UK: IEEE Computer Society) pp 843-848

Hwang B W, Blanz V, Vetter T, Lee S W, 2000 "Face reconstruction from a small number of features points", in *15th International Conference on Pattern Recognition* (Barcelona , Spain: IEEE Computer Society) pp 838-841

Jiang F, Blanz V, O'toole A J, 2006 "Probing the visual representation of faces with adaptation: A view from the other side of the mean" *Psychological Science* **17** 493-500

Jingrong J, Lijun Y, Morrissey J, 2004 "On the importance of skin color for 'other-race' effect", in *Proceedings of the 2004 IEEE International Conference on Multimedia and Expo* pp 399-402

Kassin S M, Ellsworth P C, Smith V L, 1989 "On the 'general acceptance' of eyewitness testimony research: A new survey of the experts" *American Psychologist* **44** 1089-1098

Kohonen T, 1990 "The self-organizing map" *Proceedings of the IEEE* **78** 1464-1480

Krouse F L, 1981 "Effects of pose, pose change, and delay on face recognition performance" *Journal of Applied Psychology* **66** 651-654

Lavrakas P J, Buri J R, Mayzner M S, 1976 "A perspective on the recognition of other-race faces" *Perception & Psychophysics* **20** 475-481

Lee K, Byatt G, Rhodes G, 2000 "Caricature effects, distinctiveness, and identification: Testing the face-space framework" *Psychological Science* **11** 379-385

Leopold D A, O'toole A J, Vetter T, Blanz V, 2001 "Prototype-referenced shape encoding revealed by high-level aftereffects" *Nature Neuroscience* **4** 89-94

Lewis M B, 2004 "Face-space-R: Towards a unified account of face recognition" *Visual Cognition* **11** 29-69

Light L, Kayra-Stuart F, Hollander S, 1979 "Recognition memory for typical and unusual Faces" *Journal of Experimental Psychology: Human Learning and Memory* **5** 212-228

Lindsay D S, Jack P C, Christian M A, 1991 "Other-race face perception" *Journal of Applied Psychology* **76** 587-589

Logie R H, Baddeley A D, Woodhead M M, 1987 "Face recognition, pose and ecological validity" *Applied Cognitive Psychology* **1** 53-69

Macmillan N A, Creelman C D, 2005 *Detection Theory: A User's Guide, Second Edition* (Mahwah, NJ: Lawrence Erlbaum Associates)

Malpass R S, 1981 "Training in face recognition", in *Perceiving and Remembering Faces* Eds G Davies, H Ellis and J Shepherd (London, UK: Academic Press) pp 271-285

Malpass R S, Lavigueur H, Weldon D E, 1973 "Verbal and visual training in face recognition" *Perception and Psychophysics* **14** 285-292

Mao Y, Wei H, Feng C, Sun C, 2008 "Textured 3-D face reconstruction with two images in different views", in *Proceedings of the 2008 World Automation Congress* (Waikoloa, HI: IEEE)

Marques M, Costeira J, 2008 "3D face recognition from multiple image: a shape-from-motion approach", in *8th IEEE International Conference on Automatic Face & Gesture Recognition* (Amsterdam, NL: IEEE Computer Society) pp 1-6

Matusik W, Buehler C, Raskar R, Gortler S J, McMillan L, 2000 "Image-based visual hulls", in *27th Annual Conference on Computer Graphics and Interactive Techniques* (New Orleans, LA: ACM Press/Addison-Wesley Publishing) pp 369-374

Mauro R, Kubovy M, 1992 "Caricature and face recognition" *Memory & Cognition* **20** 433-440

McCandliss B D, Fiez J A, Protopapas A, Conway M, McClelland J L, 2002 "Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception" *Cognitive, Affective, & Behavioral Neuroscience* **2** 89-108

McClelland J L, 2001 "Failures to learn and their remediation: A Hebbian account", in *Mechanisms of Cognitive Development: Behavioral and Neural Perspectives* Eds J L McClelland and R S Siegler (Mahwah, NJ: Lawrence Erlbaum Associates) pp 97-121

Meissner C A, Brigham J C, 2001 "Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review" *Psychology, Public Policy, and Law* **7** 3-35

Ming Y, Ruan Q, 2010 "3D face reconstruction using a single 2D face image", in *Proceedings of the 2010 International Conference on Educational and Information Technology* (Chongqing, China: Curran Associates) pp 32-36

Moghaddam B, Lee J, Pfister H, Machiraju R, 2003 "Model-based 3D face capture with shape-from-silhouettes", in *Proceedings of the 2003 IEEE International Workshop on*

Analysis and Modeling of Faces and Gestures (Washington, DC: IEEE Computer Society) pp 20-27

O'Toole A J, Deffenbacher K, Abdi H, Bartlett J, 1991 "Simulating the 'other-race effect' as a problem in perceptual learning" *Connection Science* **3** 163-178

O'Toole A J, Deffenbacher K A, Valentin D, Abdi H, 1994 "Structural aspects of face recognition and the other-race effect" *Memory & Cognition* **22** 208-224

O'Toole A J, Vetter T, Blanz V, 1999 "Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: An application of three-dimensional morphing" *Vision Research* **39** 3145-3155

O'Toole A J, Vetter T, Volz H, Salter E M, 1997 "Three-dimensional caricatures of human heads: distinctiveness and the perception of facial age" *Perception* **26** 719-732

O'Toole A J, Wenger M J, Townsend J T, 2001 "Quantitative models of perceiving and remembering faces: precedents and possibilities", in *Computational, Geometric, and Process Perspectives on Facial Cognition* Eds M J Wenger and J T Townsend Lawrence Erlbaum Associates) pp 1-38

Pezdek K, Blandon-Gitlin I, Moore C, 2003 "Children's face recognition memory: More evidence for the cross-race effect" *Journal of Applied Psychology* **88** 760–763

Rhodes G, Brennan S, Carey S, 1987 "Identification and ratings of caricatures: Implications for mental representation of faces" *Cognitive Psychology* **19** 473-497

Rhodes G, McLean I G, 1990 "Distinctiveness and expertise effects with homogeneous stimuli: Towards a model of configural coding" *Perception* **19** 773-794

Rhodes G, Tremewan T, 1994 "Understanding face recognition: caricature effects, inversion and the homogeneity problem" *Visual Cognition* **1** 275-311

Rodríguez J, Bortfeld H, Gutiérrez-Osuna R, 2008 "Reducing the other-race effect through caricatures", in *8th IEEE International Conference on Automatic Face and Gesture Recognition* (Amsterdam, NL: IEEE Computer Society) pp 1-6

Rodríguez J, Bortfeld H, Rudomín I, Hernández B, Gutiérrez-Osuna R, 2009 "The reverse-caricature effect revisited: Familiarization with frontal facial caricatures improves veridical face recognition" *Applied Cognitive Psychology* **23** 733-742

Rodríguez J, Gutierrez-Osuna R, 2011 "Reverse caricatures effects on three-dimensional facial reconstructions" *Image and Vision Computing* **29** 329-334

Sangrigoli S, Pallier C, Argenti A M, Ventureyra V A G, de Schonen S, 2005 "Reversibility of the other-race effect in face recognition during childhood" *Psychological Science* **16** 440-444

Shapiro P, Penrod S D, 1986 "A meta analysis of the facial identification literature" *Psychological Bulletin* **100** 139-156

Shepherd J, 1981 "Social factors in face recognition", in *Perceiving and Remembering Faces* Eds G Davies, H Ellis and J Shepherd (San Diego, CA: Academic Press) pp 55-79

Shepherd J, Deregowski J, 1981 "Races and faces: A comparison of the responses of Africans and Europeans to faces of the same and different races" *British Journal of Social Psychology* **20** 125-133

Sinha P, Balas B, Ostrovsky Y, Russell R, 2006 "Face Recognition by humans: nineteen results all computer vision researchers should know about" *Proceedings of the IEEE* **94** 1948-1962

Stevenage S V, 1995 "Can caricatures really produce distinctiveness effects?" *British Journal of Psychology* **86** 127-146

Suen C Y, Langaroudi A Z, Feng C, Mao Y, 2007 "A survey of techniques for face reconstruction", in *IEEE International Conference on Systems, Man and Cybernetics* (Montreal, Quebec: Systems, Man and Cybernetics Society) pp 3554-3560

Szeliski R, 1993 "Rapid octree construction from image sequences" *Computer Vision, Graphics, and Image Processing: Image Understanding* **58** 23-32

Tanaka J W, Simon V B, 1996 "Caricature recognition in a neural network" *Visual Cognition* **3** 305-324

Turk M, Pentland A, 1991 "Eigenfaces for recognition" *Journal of Cognitive Neuroscience* **3** 71-86

Tversky B, Baratz D, 1985 "Memory for faces: Are caricatures better than photographs?" *Memory & Cognition* **13** 45-49

Valentin D, Abdi H, Edelman B, 1997 "What represents a face: A computational approach for the integration of physiological and psychological data" *Perception* **26** 1271-1288

Valentine T, 1990 "Representation and process in face recognition" *Vision and Visual Dysfunction* **14** 107-124

Valentine T, 1991 "A unified account of the effects of distinctiveness, inversion and race in face recognition" *Quarterly Journal of Experimental Psychology* **43A** 161-204

Valentine T, Bruce V, 1986a "The effect of race, inversion and encoding activity upon face recognition" *Acta Psychologica* **61** 259-273

Valentine T, Bruce V, 1986b "The effects of distinctiveness in recognising and classifying faces" *Perception* **15** 525-535

Valentine T, Bruce V, 1986c "Recognizing familiar faces: The role of distinctiveness and familiarity" *Canadian Journal of Psychology* **40** 300-305

Valentine T, Endo M, 1992 "Towards an exemplar model of face processing: The effects of race and distinctiveness" *The Quarterly Journal of Experimental Psychology* **44** 671-703

Vrij A, Winkel F W, 1989 "Recognition of own and other-race faces: The effects of quality versus quantity of cross-cultural interaction" *Journal of Police and Criminal Psychology* **5** 24-28

Widanagamaachchi W N, Dharmaratne A T, 2008 "3D face reconstruction from 2D images: A survey", in *Digital Image Computing: Techniques and Applications* (Canberra, Australia: IEEE Computer Society) pp 365-371

Winograd E, 1981 "Elaboration and distinctiveness in memory for faces." *Journal of Experimental Psychology: Human Learning and Memory* **7** 181-190

Yip B, Jin J S, 2005 "A novel approach of 3D reconstruction of human face using monocular camera", in *11th International Multimedia Modeling Conference* (Melbourne, Australia: IEEE Computer Society) pp 378-385

Zhang R, Tsai P S, Cryer J E, Shah M, 1999 "Shape from shading: A survey" *IEEE Trans. on Pattern Analysis and Machine Intelligence* **21** 690-706

APPENDIX A

3DFS-100 DATASET

The 3DFS-100 dataset was developed at the University of Freiburg in Germany based on data captured at the University of South Florida as part of the DARPA HumanID project. The dataset contains the PCA decomposition of 100 face shapes and textures with full correspondence across faces. Face shape and texture are defined in the dataset as in equations (A.1) and (A.2), where m is the number of principal components (in this dataset is 100), α_i are the shape coefficients, s_i are shape eigenvectors, β_i are the texture coefficients, and t_i are the texture eigenvectors. Notice that both shape and texture coefficients are normalized. The standard deviation for each coefficient is provided in database.

$$S = s_{avg} + \sum_{i=1}^{m-1} \alpha_i \cdot s_i \tag{A.1}$$

$$T = t_{avg} + \sum_{i=1}^{m-1} \beta_i \cdot t_i \tag{A.2}$$

The author would like to acknowledge Professor Sudeep Sarkar at the University of South Florida for providing access to this dataset.

APPENDIX B

DERIVATIVE OF COST FUNCTION

This abstract provides a derivation of the iterative optimization algorithm described in Blanz and Vetter (2002), namely the calculation of gradient of the input image cost function E_I by α_i and β_i coefficients in equation (B.1). This derivation provides details not included in the original publication, but which were important in order to implement the method.

As described in Section 3, the cost function E_I is defined by the sum of the squared differences of the input image (target) and the image rendered from the 3D model.

$$E_I = \sum_x \sum_y \left\| I_{input}(x,y) - I_{model}(x,y) \right\|^2 \tag{B.1}$$

The input image I_{input}, and the rendered image from the 3D model I_{model}, are defined by the three color channels and pixel indices x and y.

$$I_{input}(x,y) = \left(I_r(x,y), I_g(x,y), I_b(x,y) \right)^T \tag{B.2}$$

Evaluating E_I at the center of k randomly selected triangles results in:

$$E_I = \sum_k \sum_{f=r,g,b} \left\| I_{f,input}(p_{x,k}, p_{y,k}) - I_{f,model,k} \right\|^2 \tag{B.3}$$

Notice that $p_{x,k}$ and $p_{y,k}$ are the pixel indices estimated from the perspective projection of the selected triangles from 3D model. Using the chain rule below, we can compute $\frac{\partial E_I}{\partial \alpha_i}$. We will also compute $\frac{\partial E_I}{\partial \beta_i}$ similarly later on.

$$\frac{d}{dx} f(u) = f'(u) \frac{du}{dx} \tag{B.4}$$

$$u = I_{f,input}(p_{xk}, p_{yk}) - I_{f,model,k}$$

$$f(u) = \sum \sum u^2$$

$$f'(u) = 2 \sum \sum u \cdot \frac{\partial u}{\partial \alpha_i} \tag{B.5}$$

$$\frac{\partial u}{\partial \alpha_i} = \frac{\partial}{\partial \alpha_i} I_{f,input}(p_{x,k}, p_{y,k}) - \frac{\partial}{\partial \alpha_i} I_{f,model,k}$$

$$\frac{\partial E_I}{\partial \alpha_i} = 2 \sum_k \sum_{f=r,g,b} \left(I_{f,input}(p_{x,k}, p_{y,k}) - I_{f,model,k} \right)$$

$$\cdot \left(\frac{\partial}{\partial \alpha_i} I_{f,input}(p_{x,k}, p_{y,k}) - \frac{\partial}{\partial \alpha_i} I_{f,model,k} \right) \tag{B.6}$$

$$\frac{\partial E_I}{\partial \alpha_i} = 2 \sum_k \sum_{f=r,g,b} \left(I_{f,input}(p_{x,k}, p_{y,k}) - I_{f,model,k} \right)$$

$$\cdot \left(\frac{\partial}{\partial \alpha_i} I_{f,input}(p_{x,k}, p_{y,k}) - \frac{\partial}{\partial \alpha_i} I_{f,model,k} \right) \tag{B.7}$$

From equation (B.7), we need to solve the partial derivatives: $\frac{\partial}{\partial \alpha_i} I_{f,input}(p_{x,k}, p_{y,k})$ and

$\frac{\partial}{\partial \alpha_i} I_{f,model,k}$. Using the following definition for multivariate functions,

$$\left. \begin{aligned} w &= f(u,v) \\ u &= g(x) \\ v &= k(x) \end{aligned} \right\} \quad \Rightarrow \quad \frac{\partial w}{\partial x} = \frac{\partial w}{\partial u} \cdot \frac{\partial u}{\partial x} + \frac{\partial w}{\partial v} \cdot \frac{\partial v}{\partial x} \tag{B.8}$$

we obtain:

$$\frac{\partial}{\partial \alpha_i} I_{f,input}(p_{x,k}, p_{y,k}) = \frac{\partial}{\partial x} I_{f,input} \cdot \frac{\partial}{\partial \alpha_i} p_{x,k} + \frac{\partial}{\partial y} I_{f,input} \cdot \frac{\partial}{\partial \alpha_i} p_{y,k} \tag{B.9}$$

The partial derivatives $\frac{\partial I_{f,input}}{\partial x}$ and $\frac{\partial I_{f,input}}{\partial y}$ are estimated using a discrete differentiation

operation called the Sobel operator:

$$\frac{\partial I_{f,input}}{\partial x} = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I_{f,input}$$

$$\frac{\partial I_{f,input}}{\partial y} = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I_{f,input} \tag{B.10}$$

Using perspective projections (B.11) we can solve for $\frac{\partial p_{x,k}}{\partial \alpha_i}$ and $\frac{\partial p_{y,k}}{\partial \alpha_i}$ in equation (B.9) as:

$$
\begin{aligned}
p_{x,k} &= P_x + f\left(\frac{w_{x,k}}{w_{z,k}}\right) \\
p_{y,k} &= P_y + f\left(\frac{w_{y,k}}{w_{z,k}}\right)
\end{aligned}
\tag{B.11}
$$

where (P_x, P_y) is the position of the image plane, f is the camera focal length, and the k triangles' 3D world coordinates are defined below:

$$
w_k = \left(w_{x,k}, w_{y,k}, w_{z,k}\right)^T = R_\gamma R_\theta R_\phi \hat{x}_k + t_w
\tag{B.12}
$$

where $R_\gamma R_\theta R_\phi$ are the rotation matrices, t_w is the spatial shift, and $\hat{x}_k$ are the coordinates of the center of the randomly selected k triangles. $\hat{x}_k$ is defined as:

$$
\hat{x}_k = (x_k, y_k, z_k)^T
\tag{B.13}
$$

Thus $\frac{\partial p_{x,k}}{\partial \alpha_i}$ is derived as follows:

$$
\frac{\partial p_{x,k}}{\partial \alpha_i} = \frac{\partial P_x}{\partial \alpha_i} + f\frac{\partial}{\partial \alpha_i} \cdot \frac{w_{x,k}}{w_{z,k}}
\tag{B.14}
$$

where $\frac{\partial P_x}{\partial \alpha_i} = 0$, and $f\frac{\partial}{\partial \alpha_i} \cdot \frac{w_{x,k}}{w_{z,k}}$ is solved using the quotient rule, therefore:

$$\frac{\partial p_{x,k}}{\partial \alpha_i} = f \left(\frac{\frac{\partial w_{x,k}}{\partial \alpha_i} \cdot w_{z,k} - w_{x,k} \cdot \frac{\partial w_{z,k}}{\partial \alpha_i}}{w_{z,k}^2} \right)$$

$$= f \left(\frac{\partial w_{x,k}}{\partial \alpha_i} \cdot \frac{1}{w_{z,k}} - \frac{w_{x,k}}{w_{z,k}^2} \cdot \frac{\partial w_{z,k}}{\partial \alpha_i} \right) \qquad \text{(B.15)}$$

$$= \frac{f}{w_{z,k}} \left(\frac{\partial w_{x,k}}{\partial \alpha_i} - \frac{w_{x,k}}{w_{z,k}} \cdot \frac{\partial w_{z,k}}{\partial \alpha_i} \right)$$

After some algebraic manipulations (B.16) on the perspective projection equations (B.11),

$$\frac{w_{x,k}}{w_{z,k}} = \frac{p_{xk} - P_x}{f}; \quad \frac{w_{y,k}}{w_{z,k}} = \frac{P_y - p_{y,k}}{f} \qquad \text{(B.16)}$$

$\frac{\partial p_{x,k}}{\partial \alpha_i}$ can be rewritten as:

$$\frac{\partial p_{x,k}}{\partial \alpha_i} = \frac{1}{w_{z,k}} \left(f \cdot \frac{\partial w_{x,k}}{\partial \alpha_i} - \left(p_{x,k} - P_x \right) \cdot \frac{\partial w_{z,k}}{\partial \alpha_i} \right) \qquad \text{(B.17)}$$

Similarly, $\frac{\partial p_{y,k}}{\partial \alpha_i}$ can be simplified as follows:

$$\frac{\partial p_{y,k}}{\partial \alpha_i} = \frac{-1}{w_{z,k}} \left(f \cdot \frac{\partial w_{y,k}}{\partial \alpha_i} + \left(p_{y,k} - P_y \right) \cdot \frac{\partial w_{z,k}}{\partial \alpha_i} \right) \qquad \text{(B.18)}$$

The $\frac{\partial w_{x,k}}{\partial \alpha_i}$, $\frac{\partial w_{y,k}}{\partial \alpha_i}$, and $\frac{\partial w_{z,k}}{\partial \alpha_i}$ partial derivatives from equations (B.17) and (B.18) are obtained by computing the partial derivative, $\frac{\partial w_k}{\partial \alpha_i}$, of equation (B.12):

$$\frac{\partial w_k}{\partial \alpha_i} = \frac{\partial}{\partial \alpha_i}\left(R_\gamma R_\theta R_\phi \hat{x}_k\right) + \frac{\partial}{\partial \alpha_i} t_w \tag{B.19}$$

The term $\frac{\partial}{\partial \alpha_i} t_w$ is equal to 0, therefore:

$$\frac{\partial w_k}{\partial \alpha_i} = R_\gamma R_\theta R_\phi \frac{\partial \hat{x}_k}{\partial \alpha_i} \tag{B.20}$$

The $\frac{\partial \hat{x}_k}{\partial \alpha_i}$ factor can be obtained using the shape equation (A.1) at each selected triangle:

$$\frac{\partial \hat{x}_k}{\partial \alpha_i} = \frac{\partial s_{avg}}{\partial \alpha_i} + \frac{\partial}{\partial \alpha_i}(\alpha_i s_{i,k}); \quad \frac{\partial s_{avg}}{\partial \alpha_i} = 0$$

$$\frac{\partial \hat{x}_k}{\partial \alpha_i} = s_{i,k}\big|_{k_1,k_2,k_3 \in k} = \frac{1}{3}\left(\frac{\partial x_{k_1}}{\partial \alpha_i} + \frac{\partial x_{k_2}}{\partial \alpha_i} + \frac{\partial x_{k_3}}{\partial \alpha_i}\right) \tag{B.21}$$

Now, going back to equation (B.7), we need to solve the remaining derivative, $\frac{\partial}{\partial \alpha_i} I_{f,model,k}$, to complete the $\frac{\partial E_I}{\partial \alpha_i}$ derivation. The rendered image I_f, is defined by the gain g_f, the contrast c, and the overall luminance L.

$$I_f = g_f(c \cdot L_f + (1 - c)L) + o_f$$

$$L = 0.3L_r + 0.59L_g + 0.11L_b \tag{B.22}$$

Notice that $f = r, g, b$ for each of the color channels. Using (B.22), $\frac{\partial I_f}{\partial \alpha_i}$ is expanded as:

$$\frac{\partial I_f}{\partial \alpha_i} = g_f \left(c \cdot \frac{\partial L_f}{\partial \alpha_i} + (1 - c) \left(0.3 \cdot \frac{\partial L_r}{\partial \alpha_i} + 0.59 \cdot \frac{\partial L_g}{\partial \alpha_i} + 0.11 \cdot \frac{\partial L_b}{\partial \alpha_i} \right) \right) + \frac{\partial o_f}{\partial \alpha_i} \tag{B.23}$$

The term $\frac{\partial o_f}{\partial \alpha_i}$ is equal to zero. Now, $\frac{\partial L_f}{\partial \alpha_i}$ is obtained using the Phong illumination equation below:

$$L_{f,k} = F_k \cdot L_{f,amb} + F_k \cdot L_{f,dir}\langle n_k, l \rangle + k_s \cdot L_{f,dir}\langle r_k, \hat{v}_k \rangle^v \tag{B.24}$$

Where:

- $F_k = R_k, G_k, B_k$ (each color channel)

- $f = r, g, b$

- $l = [\cos(\theta_l) \cdot \sin(\phi_l), \ \sin(\theta_l), \ \cos(\theta_l)\cos(\phi_l)]^T$ defines the light direction.

- n_k is normal vector of the kth triangle.

- $\hat{v}_k$ is the viewing direction.

- r_k is the direction of the maximum specular reflection.

- k_s is the specular reflectance.

- v is the angular distribution of the specular reflection.

Then, $\frac{\partial L_f}{\partial \alpha_i}$ is written as:

$$\frac{\partial L_f}{\partial \alpha_i} = \frac{\partial}{\partial \alpha_i}\left(F_k \cdot L_{f,amb}\right) + \frac{\partial}{\partial \alpha_i}\left(F_k \cdot L_{f,dir} \cdot \langle n_k, l\rangle\right) + \frac{\partial}{\partial \alpha_i}\left(k_s \cdot L_{f,dir}\langle r_k, \hat{v}_k\rangle^\nu\right) \qquad \text{(B.25)}$$

The term $\frac{\partial}{\partial \alpha_i}\left(F_k \cdot L_{f,amb}\right) = 0$, then

$$\frac{\partial L_f}{\partial \alpha_i} = F_k \cdot L_{f,dir} \cdot \langle \frac{\partial n_k}{\partial \alpha_i}, l\rangle + k_s \cdot L_{f,dir} \cdot \nu \cdot \langle r_k, \hat{v}_k\rangle^{\nu-1} \cdot \langle \frac{\partial r_k}{\partial \alpha_i}, \hat{v}_k\rangle \qquad \text{(B.26)}$$

r_k is defined as $2 \cdot \langle n_k, l\rangle \cdot n_k + l$, therefore:

$$\frac{\partial r_k}{\partial \alpha_i} = 2 \cdot \frac{\partial}{\partial \alpha_i}\left(\langle n_k, l\rangle \cdot n_k\right) + \frac{\partial l}{\partial \alpha_i} \qquad \text{(B.27)}$$

The term $\frac{\partial l}{\partial \alpha_i} = 0$, and after applying the product rule, $\frac{\partial r_k}{\partial \alpha_i}$ is defined as:

$$\frac{\partial r_k}{\partial \alpha_i} = 2 \cdot \left(\langle \frac{\partial n_k}{\partial \alpha_i}, l\rangle \cdot n_k + \langle n_k, l\rangle \cdot \frac{\partial n_k}{\partial \alpha_i}\right) \qquad \text{(B.28)}$$

Now we need to compute $\frac{\partial n_k}{\partial \alpha_i}$ term in equation (B.28). Here, $n_k = R_\gamma R_\theta R_\phi \hat{n}_k$ where $\hat{n}_k = \frac{\underline{n}_k}{\|\underline{n}_k\|}$ and $\underline{n}_k = \left(x_{k_1} - x_{k_2}\right) \times \left(x_{k_1} - x_{k_3}\right)$, therefore:

$$\frac{\partial \underline{n}_k}{\partial \alpha_i} = \frac{\partial}{\partial \alpha_i} \left(\left(x_{k_1} - x_{k_2} \right) \times \left(x_{k_1} - x_{k_3} \right) \right) \tag{B.29}$$

The above partial derivative is solved using the following definition:

$$\frac{d}{dx}(r \times u) = \left(r \times \frac{du}{dx} \right) + \left(\frac{dr}{dx} \times u \right) \tag{B.30}$$

Therefore:

$$\frac{d\underline{n}_k}{d\alpha_i} = \left(x_{k_1} - x_{k_2} \right) \times \left(\frac{\partial x_{k_1}}{\partial \alpha_i} - \frac{\partial x_{k_3}}{\partial \alpha_i} \right) + \left(\frac{\partial x_{k_1}}{\partial \alpha_i} - \frac{\partial x_{k_2}}{\partial \alpha_i} \right) \times \left(x_{k_1} - x_{k_3} \right) \tag{B.31}$$

Now, we compute $\frac{\partial \hat{n}_k}{\partial \alpha_i}$ as follows:

$$\frac{\partial \hat{n}_k}{\partial \alpha_i} = \frac{\partial}{\partial \alpha_i} \left(\frac{n_k}{\|\underline{n}_k\|} \right) = \frac{\frac{\partial \underline{n}_k}{\partial \alpha_i} \cdot \|\underline{n}_k\| - \frac{\partial \|\underline{n}_k\|}{\partial \alpha_i} \cdot \underline{n}_k}{\|\underline{n}_k\|^2} \tag{B.32}$$

Using the following definition:

$$\frac{d}{dx} \|f(x)\| = \frac{\langle \frac{df(x)}{dx}, f(x) \rangle}{\|f(x)\|} \tag{B.33}$$

$\frac{\partial \hat{n}_k}{\partial \alpha_i}$ is rewritten as follows:

$$\frac{\partial \hat{n}_k}{\partial \alpha_i} = \left\| \underline{n}_k \right\|^{-1} \cdot \frac{\partial \underline{n}_k}{\partial \alpha_i} - \frac{\langle \frac{\partial \underline{n}_k}{\partial \alpha_i}, \underline{n}_k \rangle \cdot \underline{n}_k}{\left\| \underline{n}_k \right\|^3} \tag{B.34}$$

After some algebraic manipulation, and given that $\hat{n}_k = \frac{n_k}{\left\| \underline{n}_k \right\|}$, we finally obtained:

$$\frac{\partial \hat{n}_k}{\partial \alpha_i} = \left\| \underline{n}_k \right\|^{-1} \cdot \left(\frac{\partial \underline{n}_k}{\partial \alpha_i} - \langle \frac{\partial \underline{n}_k}{\partial \alpha_i}, \hat{n}_k \rangle \cdot \hat{n}_k \right) \tag{B.35}$$

Recalling that $n_k = R_\gamma R_\theta R_\phi \hat{n}_k$, therefore:

$$\frac{\partial n_k}{\partial \alpha_i} = R_\gamma R_\theta R_\phi \cdot \frac{d \hat{n}_k}{d \alpha_i} \tag{B.36}$$

This completes $\frac{\partial E_I}{\partial \alpha_i}$ derivation. Now, the partial derivative $\frac{\partial E_I}{\partial \beta_i}$ is computed in similar fashion using (B.5):

$$\begin{aligned} \frac{\partial E_I}{\partial \beta_i} = 2 \sum_k \sum_{f=r,g,b} & \left(I_{f,input}(p_{x,k}, p_{y,k}) - I_{f,model,k} \right) \\ & \cdot \left(\frac{\partial}{\partial \beta_i} I_{f,input}(p_{x,k}, p_{y,k}) - \frac{\partial}{\partial \beta_i} I_{f,model,k} \right) \end{aligned} \tag{B.37}$$

but this time $\frac{\partial}{\partial \beta_i} I_{f,input}(p_{x,k}, p_{y,k}) = 0$, therefore:

$$\frac{\partial E_I}{\partial \beta_i} = -2 \sum_k \sum_{f=r,g,b} \left(I_{f,input}(p_{x,k}, p_{y,k}) - I_{f,model,k} \right) \cdot \left(\frac{\partial}{\partial \beta_i} I_{f,model,k} \right). \tag{B.38}$$

The only derivative that remains to solve is $\frac{\partial}{\partial \beta_i} I_{f,model,k}$. Using the color contrast equation (B.22),

$$\frac{\partial}{\partial \beta_i} I_{f,model,k} = g_f \left(c \cdot \frac{\partial L_f}{\partial \beta_i} + (1-c) \left(0.3 \cdot \frac{\partial L_r}{\partial \alpha_i} + 0.59 \cdot \frac{\partial L_g}{\partial \alpha_i} + 0.11 \cdot \frac{\partial L_b}{\partial \alpha_i} \right) \right) \tag{B.39}$$

$\frac{\partial L_f}{\partial \beta_i}$ is calculated using Phong's illumination equation:

$$\frac{\partial L_f}{\partial \beta_i} = \frac{\partial}{\partial \beta_i} \left(F_k \cdot L_{f,amb} + F_k \cdot L_{f,dir} \langle n_k, l \rangle + k_s \cdot L_{f,dir} \langle r_k, \hat{v}_k \rangle^v \right)$$

$$\frac{\partial L_f}{\partial \beta_i} = \frac{\partial F_k}{\partial \beta_i} \left(L_{f,amb} + L_{f,dir} \langle n_k, l \rangle \right) + \frac{\partial}{\partial \beta_i} \cdot \left(k_s \cdot L_{f,dir} \langle r_k, \hat{v}_k \rangle^v \right) \tag{B.40}$$

The term $\frac{\partial}{\partial \beta_i} \cdot \left(k_s \cdot L_{f,dir} \langle r_k, \hat{v}_k \rangle^v \right) = 0$ and $\frac{\partial F_k}{\partial \beta_i}$ can be obtained from equation (A.2):

$$\frac{\partial F_k}{\partial \beta_i} = \frac{\partial t_{avg}}{\partial \beta_i} + \frac{\partial}{\partial \beta_i} \cdot \beta_i t_{i,k} = t_{i,k} \tag{B.41}$$

$$\frac{\partial F_k}{\partial \beta_i} = t_{i,k}|_{k1,k2,k3 \in k} = \frac{1}{3} \left(\frac{\partial F_{k1}}{\partial \beta_i} + \frac{\partial F_{k2}}{\partial \beta_i} + \frac{\partial F_{k3}}{\partial \beta_i} \right) \tag{B.42}$$

therefore:

$$\frac{\partial L_f}{\partial \beta_i} = t_{i,k} \cdot \left(L_{f,amb} + L_{f,dir} \langle n_k, l \rangle \right) \tag{B.43}$$

and we are done.

APPENDIX C

MORPHABLE MODEL USER'S GUIDE

System Requirements

Our implementation of the morphable model software has been tested on a Pentium® 4

with 2GHz and 2.5 GB of memory. Table C.1 shows additional system requirements and

dependencies.

Table C.1. System requirement dependencies.

Component	Notes
Video Card	The morphable model implementation requires a video card supporting OpenGL® 2.0. The current implementation was tested on *n*VIDIA® 6800 GS and ATI Radeon™ X1600 PRO cards.
Microsoft® Visual Studio®	OpenGL tools and libraries, and MATLAB® MEX functions were implemented and tested using Visual Studio® 2005 environment.
Mathworks MATLAB®	The morphable model optimization code was implemented mostly in MATLAB. The current implementation has been tested in MATLAB 2006a and 2007a versions.
ActiveTcl	Required for the distribution generation scripts. Version used was 8.4.13.0.
GLUT library	This is the OpenGL Utility Toolkit required by the rendering tools developed as part of this distribution. GLUT is a programming interface with ANSI C and FORTRAN binding for writing windows independent OpenGL programs.
3DFS dataset	Dataset containing the PCA decomposition of 100 face shapes and textures with full correspondence across faces.

Installation and build

The following subsections describe the procedures to load the morphable model software into your system. Additionally, it shows how to rebuild some of the software components if needed. Finally, it shows how to remove the software.

Loading source and runtime distribution

Insert DVD on your computer and using the MATLAB environment run 'installation.m' script located in the 'install' directory. The following shows a sample output from the installation script assuming the DVD unit is in 'D:' drive.

```
>> cd d:\install
>> installation

----------------------------------------------------------------
          MORPHABLE SOFTWARE INSTALLATION SCRIPT v1.00
----------------------------------------------------------------

[CREATING TARGET]
Creating C:\Morphable...
Directory C:\Morphable created...

[INSTALLING CORE MODULES]
Creating C:\Morphable\matlab...
Directory C:\Morphable\matlab created...
Copying facefit into C:\Morphable\matlab...
Copying html into C:\Morphable\matlab...
Copying mat into C:\Morphable\matlab...
Copying optimizer into C:\Morphable\matlab...
Copying segmentation into C:\Morphable\matlab...
Copying utils into C:\Morphable\matlab...
Creating C:\Morphable\vstudio...
Directory C:\Morphable\vstudio created...
Copying 3dFaces.ncb into C:\Morphable\vstudio...
Copying 3dFaces.sln into C:\Morphable\vstudio...
Copying FaceRendering into C:\Morphable\vstudio...
Copying MatrixLib into C:\Morphable\vstudio...
Copying PcLoader into C:\Morphable\vstudio...
Copying SharedMemory into C:\Morphable\vstudio...
Copying facedb into C:\Morphable\vstudio...
Copying multiprod into C:\Morphable\vstudio...
Copying readexr into C:\Morphable\vstudio...
Copying renderface into C:\Morphable\vstudio...
```

```
Copying writeexr into C:\Morphable\vstudio...

[CONFIGURING CORE MODULES]
Setting up MEX files...
Saving data files...
Setting up paths...


------------------------------------------------------------
Done!
>>
```

The installation script places the entire distribution in 'C:\Morphable'

directory and adds the necessary search paths in MATLAB environment.

Rebuilding C/C++ software (if needed)

This distribution also contains the C/C++ source code for the custom MEX

functions used the morphable model software. If you need to recompile them

execute the commands below in the Visual Studio 2005 command prompt.

Please notice, this distribution comes with MEX DLLs already compiled. These

instructions will be required only if a new version of MATLAB is not compatible

with the installed MEX functions.

```
C:\> cd Morphable\vstudio
C:\Morphable\vstudio> devenv 3dFaces.sln /build

Microsoft (R) Visual Studio Version 8.0.50727.762.
Copyright (C) Microsoft Corp 1984-2005. All rights reserved.

-- Build started: Project: MatrixLib, Configuration: Debug Win32 --
     Creating library...
     Build log was saved at
     "file://c:\Morphable\vstudio\MatrixLib\Debug\BuildLog.htm"
     MatrixLib - 0 error(s), 0 warning(s)

-- Build started: Project: PcLoader, Configuration: Debug Win32 --
     Creating library...
     Build log was saved at
     "file://c:\Morphable\vstudio\PcLoader\Debug\BuildLog.htm"
     PcLoader - 0 error(s), 0 warning(s)

-- Build started: Project: multiprod, Configuration: Debug Win32 --
     Linking...
```

```
        Creating library Debug/multiprod.lib and object
        Debug/multiprod.exp
        Embedding manifest...
        Build log was saved at
        "file://c:\Morphable\vstudio\multiprod\Debug\BuildLog.htm"
        multiprod - 0 error(s), 0 warning(s)

-- Build started: Project: renderface, Configuration: Debug Win32 --
        Linking...
        Creating library Debug/renderface.lib and object
        Debug/renderface.exp
        Embedding manifest...
        Build log was saved at
        "file://c:\Morphable\vstudio\renderface\Debug\BuildLog.htm"
        renderface - 0 error(s), 0 warning(s)

-- Build started: Project: SharedMemory, Configuration: Debug Win32 --
        Linking...
        Embedding manifest...
        Build log was saved at
        "file://c:\Morphable\vstudio\SharedMemory\Debug\BuildLog.htm"
        SharedMemory - 0 error(s), 0 warning(s)

-- Build started: Project: FaceRendering, Configuration: Debug Win32 --
        Linking...
        Embedding manifest...
        Build log was saved at
        "file://c:\Morphable\vstudio\FaceRendering\Debug\BuildLog.htm"
        FaceRendering - 0 error(s), 0 warning(s)

== Build: 6 succeeded, 0 failed, 0 up-to-date, 0 skipped ==

C:\Morphable\vstudio>
```

You must copy the new 'renderface.dll' and 'multiprod.dll'
to 'C:\Morphable\matlab\mex' directory if they were recompiled.

Removing software distribution

Insert DVD on your computer and using the MATLAB environment run
'cleanup.m' script located in the 'install' directory. Please close all
windows and programs accessing any directory or file on 'C:\Morphable'.
The following shows a sample output from the cleanup script assuming the
DVD unit is in 'D:' drive.

```
>> cleanup

--------------------------------------------------------------
          MORPHABLE SOFTWARE CLEANUP SCRIPT v1.00
--------------------------------------------------------------

[UNLOADING MEX FILES]
Unloading multiprod...
Unloading renderface...

[REMOVING MATLAB PATHS]
Removing C:\Morphable\matlab\facefit...
Removing C:\Morphable\matlab\mat...
Removing C:\Morphable\matlab\mex...
Removing C:\Morphable\matlab\optimizer...
Removing C:\Morphable\matlab\segmentation...
Removing C:\Morphable\matlab\utils...

[REMOVING MORPHABLE SOFTWARE]
Deleting C:\Morphable...

--------------------------------------------------------------
Done!

>>
```

High level software description

The morphable model software implements part of methodology described in Blanz
and Vetter (2002) and (2003) to estimate a 3D model from a single frontal image. The
following subsections provide a brief description of some of the main functions and
executables of the morphable model software. Detailed information is provided in the
online documentation.

Optimizer

Table C.2. Optimizer module core functions. This module contains additional functions not listed in this table; refer to the online documentation for a full list of functions.

Function	Description
derivativeCostFunction	Computes the derivative of the cost function.
init	Morphable system and rendering parameters initialization.
optimizer	Kick off the optimization procedure.
updateHessian	Re-computes the Hesssian matrix using numerical differentiation.

Face fitting

Table C.3. The face fitting module is available as prototype. Face fitting module allows the user to select 5 features from the target image, and uses those features to fit the average 3D model on the target image.

Parameter	Description
testoptim	Driver test script used to prototype the face fitting functionality. Use this script
facefit	Function to be used in the morphable module implementation. This function is based on 'testoptim' script. Integration still under development.
getFaceFeatures	Implements manual face feature selection. It returns the row and column coordinates of the selected five features.
featuresOptim	Implements the cost function used by MATLAB Optimization Toolbox.

Segmentation

This module is currently under development.

Utils

Table C.4. Utility functions. The main utility functions deal with rendering and plots, and to load the 3DFS dataset into MATLAB. Additional functions are described in the online documentation.

Parameter	Description
`buildMorphableModelDb`	Builds 3DFS face database and saves it as .mat format.
`getFace`	Projects a face and renders it using OpenGL.
`getImage`	Generates an image from a OpenGL rendered face.
`glplot`	Displays an OpenGL rendering.

Data sources

Table C.5. MATLAB data files used by the morphable model distribution.

Data files	Description
`morphable.mat`	The 3DFS dataset in a MATLAB data format.
`targetXX.mat`	Target images used for testing. 'XX' is the face id. The only face ids provided are 29, 40, and 90.
`fixedXX.mat`	Data file containing a list of 40 fixed triangles. This file is only used for debugging the morphable model with random triangle selection disabled. 'XX' is the face id. The only face ids provided are 29, 40, and 90.

Online documentation

This distribution provides online documentation for the MATLAB functions. To access it, use a web browser and open `C:\Morphable\matlab\html\index.html`. You can also use MALTAB's `help` command to access each function description.

Tutorial

This section describes a walkthrough highlighting some of the functionalities of our morphable model implementation. It briefly describes the initialization process, how to estimate a 3D model from an image, and finally how to run the face fitting prototype.

Initialization

> Before kicking off the optimization procedure, one must run the initialization script to set system and rendering parameters, and load up the 3DFS dataset. Run `init.m` which is located in `C:\Morphable\matlab\optimizer` on the MATLAB environment. The script takes a few minutes and should return to prompt without feedback. Several variables will be created in the workspace. The script is preconfigured to with a target image loaded in variable 'I'. Figure shows this target image.

Figure C.1. Target Image preloaded by the initialization script. This image will be passed to the optimizer to perform a 3D face reconstruction. This face corresponds to face id #40 in the 3DFS dataset.

There are two additional target images that come with the software, which correspond to face id #29 and #90. One can temporarily change the id by setting the 'faceId' variable in the MATLAB workspace or permanently on 'init.m' to the appropriate number. Please notice that the face id is simply the index in MATLAB corresponding to a particular sample in the 3DFS dataset. Also notice that #40 corresponds to a Caucasian male, #29 is a Caucasian female, and #90 is a East Asian male.

Figure C.2. 3D reconstruction obtained at the 200,000th iteration. This image was generated using 'glplot' function which is part of 'utils' module.

Optimization

After successfully running the initialization script, start the optimization process with a maximum number of iterations set to 1[16] by executing the following MATLAB command:

```
[a b c sa sb] = optimizer(parameters, model, I,
zeros(99,1), zeros(99,1), 0, 1);
```

_____

[16] We are setting the number of iteration to 1 just to test the installation and initialization.

where 'a' is the estimated alpha coefficients, 'b' the estimated beta coefficients, 'c' is the cost history, 'sa' is the history of the alpha parameters, and 'sb' is the history of the beta coefficients. Refer to the online documentation for more details on the input parameters of the `optimizer` function.

Figure C.2 shows a 3D reconstruction of the target image (see Figure C.1) when the `optimizer` function was set to run for 200,000 iterations. The 3D rendering was generated using the following commands:

```
rFace = getFace(sa(:,200000), ...
                model.shapeComponents, ...
                model.shapeStd, ...
                model.shapeMean, ...
                sb(:,200000), ...
                model.textureComponents, ...
                model.textureStd, ...
                model.textureMean, ...
                model.triangles, ...
                model.adjacencies);

glplot(rFace);
```

Face fitting

This prototype module is not fully integrated with the morphable software, but it can be executed as a stand-alone application. This module contains a test script that exercises this prototype functionality. To start up this test script, execute the following command:
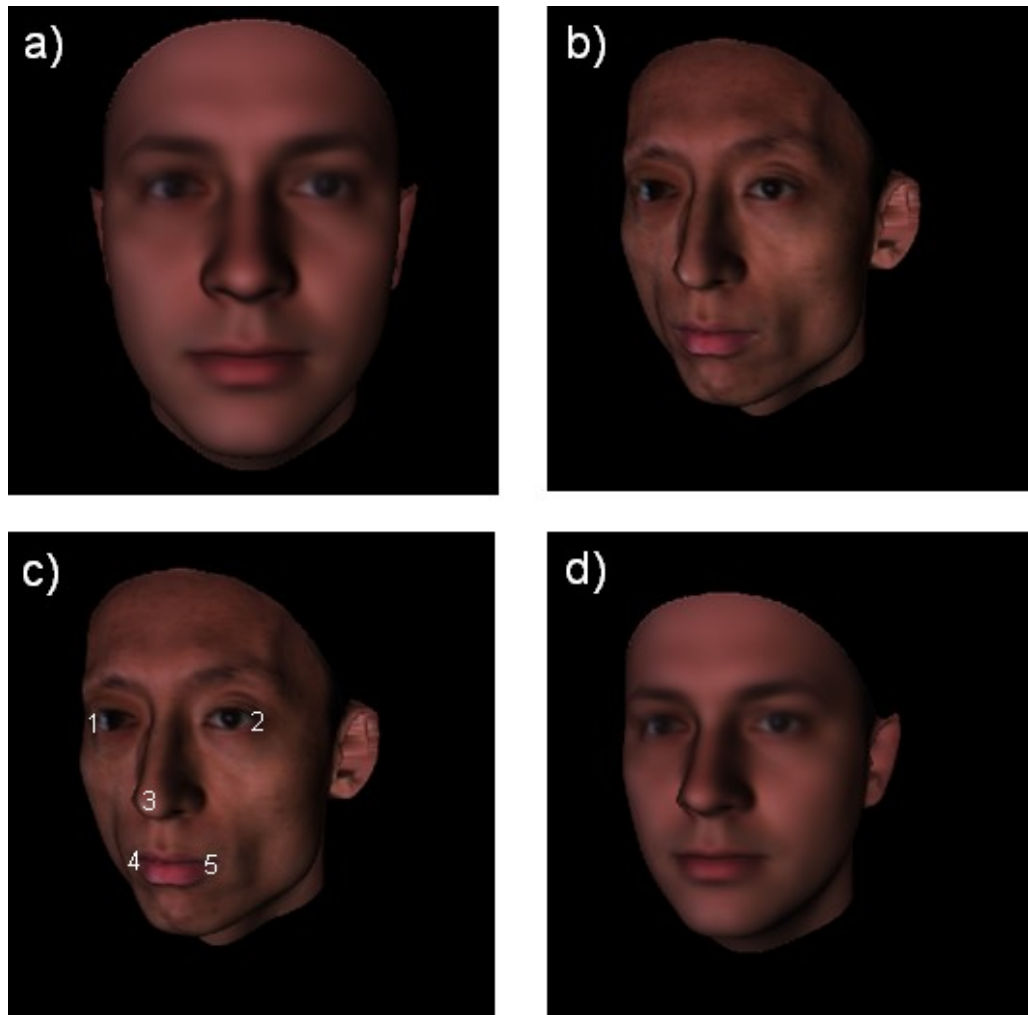
```
testoptim
```

Figure C.3. Face fitting prototype. a) shows the initial rotation and spatial shift of the average 3D face model. b) shows the target image. c) shows the location and the sequence for the facial feature selection. Finally, d) shows the result image after fitting the 3D model by optimizing the rigid transformations.

The script displays a target image (Figure C.3b) to which we want to fit the average 3D face model (Figure C.3a). On the target image, select 5 features by pressing a single left click on the following facial point in sequence: (*i*) outside corner of the model's right eye, (*ii*) outside corner of the left eye, (*iii*) the tip of the nose, (*iv*) the mouth right corner, and (*v*) the mouth left corner (Figure C.3c).

Finally, the optimizer renders the resulting model (Figure C.3d). The current idea is to use this face fitting module implementation in the morphable model's `optimizer` function. The optimization process could schedule a fitting process after a configurable number of iterations.

VITA

Jobany Rodríguez received the B.S. degree in computer engineering from the University of Puerto Rico, Mayagüez, Puerto Rico in 1995. He received the M.S. degree in electrical engineering from Monmouth University, West Long Branch, New Jersey in 1998. He also received the Ph.D. degree in computer engineering from Texas A&M University, College Station, Texas in 2011. Since 1995, he has worked as a software engineer for AT&T, Lucent Technologies, Avaya, and Honeywell Aerospace.

Jobany Rodríguez may be reached at Department of Computer Science and Engineering, Texas A&M University, TAMU 3112, College Station, TX 77843-3112. His email is jobany@hotmail.com.