Image and Vision Computing 29 (2011) 329–334

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Discussion

Reverse caricatures effects on three-dimensional facial reconstructions ☆

Jobany Rodriguez, Ricardo Gutierrez-Osuna *

Department of Computer Science and Engineering, Texas A&M University, TAMU 3112, College Station, TX 77843-3112, United States

ARTICLE INFO

Article history:
Received 9 April 2010
Received in revised form 30 October 2010
Accepted 3 January 2011

Keywords:
Face recognition
Face perception
Face reconstruction

ABSTRACT

Previous research has shown that familiarization with three-dimensional (3D) caricatures can help improve recognition of same-race and other-race faces, a result that may lead to new training tools in security applications. Since 3D facial scans are not generally available, here we sought to determine whether 3D reconstructions from 2D frontal images could be used for the same purpose. Our results suggest that, despite the high level of photographic realism achieved by current 3D facial reconstruction methods, additional research is needed in order to reduce reconstruction errors and capture the distinctive facial traits of an individual.

1. Introduction

We recognize faces from our own race better than those from another race [1–4]. This *other-race effect* (ORE) is robust, and has been reproduced in many studies [see 4–6] and across racial groups [6–8]. It is generally agreed that OREs result from the fact that the most appropriate features for discriminating faces are race-dependent [9]. For instance, 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 [10]. Research shows that this ORE may be reduced by drawing attention to the most distinctive feature of a given face. For example, Hills and Lewis [11] showed that OREs could be reduced by familiarizing subjects with own-race faces containing features critical for differentiating other-race faces.

At the same time, people have better recollection for visually distinctive faces [12–17], an effect that 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 perceptual result is known as the *caricature effect* [18–21]. Additional studies have also demonstrated a *reverse-caricature effect*, according to which familiarization with caricatures improves the recognition of the veridical face at a later time [e.g., 18–24]. These results suggest ways in which caricatures may be used as training tools in applied face recognition settings (i.e. law enforcement[1]).

Motivated by this research, our previous work [27] has explored the use of three-dimensional (3D) caricatures to direct attention to critical features of other-race faces. Our experimental results showed that reverse-caricatures reduce OREs in Caucasian participants when viewing Indian faces. Although these results are a step towards designing real-life training systems, obtaining 3D models of individuals is cost prohibitive if not impossible in some applied settings; 3D scanners are still expensive instruments, and scanning is not possible if the target individual (i.e. a crime suspect) is at large. One potential solution to this problem is to use photogrammetric techniques to reconstruct 3D face models from 2D photographs [28,29]. Using these reconstructed 3D models one could then generate caricatures from individual mug shots. However, it is unclear whether caricatures based on reconstructed 3D models are still effective, since the caricaturization process may amplify reconstruction errors to the point of rendering the caricatures unusable. Answering this question is the main objective of this work.

2. Facial reconstruction and caricaturization

For this study, we used the University of Freiburg 3DFS-100 dataset [28] containing m = 100 3D face models. Each face consisted of a mesh with n = 75,972 vertices in full correspondence, and each vertex was defined by its position in 3D Cartesian coordinates $S=(X_1,Y_1,Z_1,X_2,\ldots,Y_n,Z_n)$ and its texture in RGB space $T=(R_1,G_1,B_1,R_2,\ldots,G_n,B_n)$. Performing principal component analysis [30], a face shape and texture can be defined by:

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

$$T = t_{avg} + \sum_i^{m-1} \beta_i \cdot t_i \tag{2}$$

☆ This paper has been recommended for acceptance by Lijun Yin.

* Corresponding author. Tel.: +1 979 845 2942; fax: +1 979 847 8578.
E-mail addresses: jobany@cse.tamu.edu (J. Rodriguez), rgutier@cse.tamu.edu (R. Gutierrez-Osuna).

[1] Caricatures also have been studied as a mechanism to improve recognition of criminal facial sketches or composites [25,26].

0262-8856/$ – see front matter
doi:10.1016/j.imavis.2011.01.002

where s_{avg} and t_{avg} are the shape and texture averages, α_i are the shape principal components, s_i are the shape eigenvectors, β_i are the texture principal components and t_i are the texture eigenvectors; examples for same-race and other-race faces from the dataset are illustrated in Fig. 1a.

To test whether 3D reconstructions are amenable to reverse-caricature effects, we decided to use a best-case reconstruction scenario. Namely, we reconstructed the shape and texture of each face in the 3DFS-100 dataset in a leave-one-out fashion while holding constant the rendering parameters (i.e., camera position and illumination). First, we removed each face from the dataset and obtained a PCA decomposition on the remaining m = 99 training faces according to their shape (S_{train}) and texture (T_{train}). Next, we projected the left-out test face f_{test} along the PCA eigenvectors (s_i, t_i) to obtain its principal components α_{test} and β_{test}:

$$\alpha_{test} = \left(S_{test} - s_{avg}\right)^{\mathrm{T}} \cdot s_i \tag{3}$$

$$\beta_{test} = \left(T_{test} - t_{avg}\right)^{\mathrm{T}} \cdot t_i \tag{4}$$

which yield a reconstructed 3D model f'_{test} :

$$S'_{test} = s_{avg} + \sum_i^{m-1} \alpha_{test} \cdot s_i \tag{5}$$

$$T'_{test} = t_{avg} + \sum_i^{m-1} \beta_{test} \cdot t_i \tag{6}$$

Fig. 1. 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.

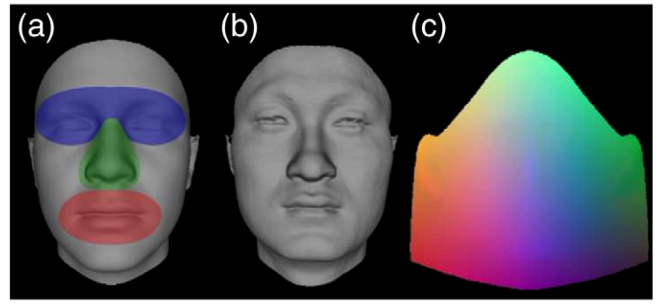

Fig. 2. (a) Face segmentation used for 3D face reconstructions; each region is predicted independently and then merged into a composite face. (b) Example of a 3D shape and (c) its corresponding geometry image; the geometry image is an $n \times m$ matrix where XYZ coordinates are represented as RGB values.

2.1. Blending 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 [28], and performed the PCA decomposition for each segment independently; see Fig. 2a. The final face model f'_{test} was obtained by combining each predicted segment through an image blending procedure [31]. Namely, given two input images (A and B) to be blended, we define a mask image M per segment that denotes whether the corresponding pixel should come from image A ($M_{ij} = 1$) or B ($M_{ij} = 0$). Then we construct a Laplace pyramid for images A and B, and a Gaussian pyramid for the mask image M, as illustrated in Fig. 3. At each level in the pyramid, the algorithm blends the two images as:

$$LC_n(i,j) = GM_n(i,j) \cdot LA_n(i,j) + (1 - GM_n(i,j)) \cdot LB_n(i,j) \tag{7}$$

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 , is synthesized from the LC_n pyramid as:

$$G_n = LC_n \tag{8}$$

$$G_{n-1} = LC_{n-1} + \mathrm{Expand}(G_n) \tag{9}$$

In order to apply this image-based blending algorithm to 3D models, the 3D segments are converted into geometry images (GI) [32] prior to the blending stage; see Fig. 2b,c. After blending all GI-based segments, the resulting GI is converted back into a 3D model. The overall method produces seamless and photorealistic reconstructions that are comparable to those in previous work [28,29,33]; examples for same-race and other-race faces are illustrated in Fig. 1b.

2.2. Caricaturization

In order to caricaturize faces consistently and evenly,[2] we first normalize each face f by its Mahalanobis distance ($\|\ \|_{\mathrm{M}}$) to the average face f_{AVG} [35]:

$$f_N = \frac{f}{\|f - f_{AVG}\|_M} \tag{10}$$

and then caricaturize it by linearly exaggerating differences with respect to f_{AVG} [19,21,35]:

$$f_C = f_{AVG} + (1 + \alpha)(f_N - f_{AVG}) \tag{11}$$

[2] Distinctive faces need to be caricaturized less than typical faces in order to achieve the same level of distinctiveness [34].