Using Virtual Heads for Person Identification:

An Empirical Study Comparing Photographs to Photogrammetrically-Generated Models

Jeremy N. Bailenson

Andrew C. Beall

Jim Blascovich

Research Center for Virtual Environments and Behavior

University of California, Santa Barbara

Jeremy N. Bailenson

Dept. of Psychology

University of California

Santa Barbara, CA 93101

bailenson@psych.ucsb.edu

Abstract

The purpose of these studies was to examine the effectiveness of virtual heads: three-dimensional models of human heads and faces. Our goal is to use these virtual head models as functional substitutes for photographs of humans as well as for live humans during eyewitness lineups and other processes relating to person identification. We tested the effectiveness of virtual heads by taking photographs of people and then using 3dMeNow software by bioVirtual to build 3D models that resembled the photographs. We tested to see how easily experimental subjects would recognize images of the 3D models (compared to photographs) after being trained on short video clips of people. The goal of this study was to examine subjects' recognition of virtual faces and to compare this performance to recognition of real faces (i.e., Troje & Bueltoff, 1996). In the following sections we discuss relevant previous research, present the methods and results of the current study, and finally point to directions for future work.

Previous Research

Previous research has studied the development of virtual heads (Liu, Collin, & Chaudhuri, 2000; Blanz & Vetter, 1999; Decarlo, Metaxas, & Stone, 1998; Brunelli, Falavigna, Poggio, & Stringa, 1995; Busey, 1988). In our previous research, we discussed the utility of using virtual reality simulations as an aid to identifying criminal suspects as well as conducting more effective police lineups (Bailenson, Wiggins, Blascovich, & Beall, 2003). Virtual reality simulations have the distinct benefit of allowing extremely realistic simulations in which the forensic specialist maintains complete control over the simulation details. For example, in reconstructing a crime scene, one can easily build an entire city block or a crowded restaurant without having to employ a contractor and construction team or hiring actors. Similarly, using virtual heads for conducting police lineups (as opposed to either live lineups or photograph lineups) has three major advantages. First, the witness or identifier has the opportunity to explore the suspect at any possible viewing angle or distance while maintaining complete anonyminity. Second, by using morphing techniques, the designer of the lineup can easily manufacture an infinite set of distracter heads (i.e., the virtual foils in the lineup) that are dissimilar to the suspect's head in terms along quantitative dimensions. Third, using animation sequences, the lineup conductor can render any facial expression or verbal utterance from the virtual head; consequently it is possible to conduct more realistic lineups than mere photographs.

In previous research, we examined the utility of using virtual heads for person identification (Bailenson, Beall, Blascovich, & Rex, 2003). More specifically, we developed a number of photogrammetrically-generated, three-dimensional, virtual heads and tested their effectiveness in terms of recognizing previously known faces in a series of 11 studies. In those studies, we employed experimental subjects and trained them to learn a series of human faces by looking at photographs. We then tested subjects' ability to recognize the faces they had learned by giving them lineups consisting of either photographs or virtual heads.

The data from those studies demonstrated an approximate seven percent disparity in recognition between photographs and virtual faces, in that recognition was facilitated better from the photographs. However, in that study, the targets (i.e., the correct responses: the faces in the lineup that they had seen previously) in the photograph condition were the exact same pictures on which the subjects were trained. This is problematic, because it could be the case that in some conditions, recognition was better in the photograph lineup than the virtual head lineup not because of any difference in quality between photographs and virtual heads, but because there were peripheral details in the photographs to which subjects were paying attention. In the new study reported in the current paper, we remedied this potential problem by training subjects on video clips of suspects moving their heads. We then conducted lineups from either photographs or from virtual heads to examine the effectiveness of virtual heads.

Current Study

In the current study, subjects were trained on video clips of human faces in the training session. After training, we tested subjects on the photographs and virtual versions of those human faces over a number of viewing angles in the testing session. Subjects were given unlimited time during the training session and the testing session to view the stimuli.

Methods. There were 26 subjects in the study who received partial course credit for an introductory psychology course for participating. Subjects were tested one at a time. In the training session, we presented them videos of 5 faces that were randomly selected from our database. Each training face was a looping, black and white, 20 second video clip of a person moving their head slowly from side to side and then up and down. We randomly paired each face with an arbitrary first name. Subjects scrolled through each video, one at a time, until they had learned the names of the set of 5 faces to a criterion of 80 percent correct.

After the training session, subjects began the testing session that gauged their recognition of the faces. We manipulated two within-subject variables, face type (photograph or virtual) and face viewing angle (Frontal, ¾ View, or Profile). Figure 1

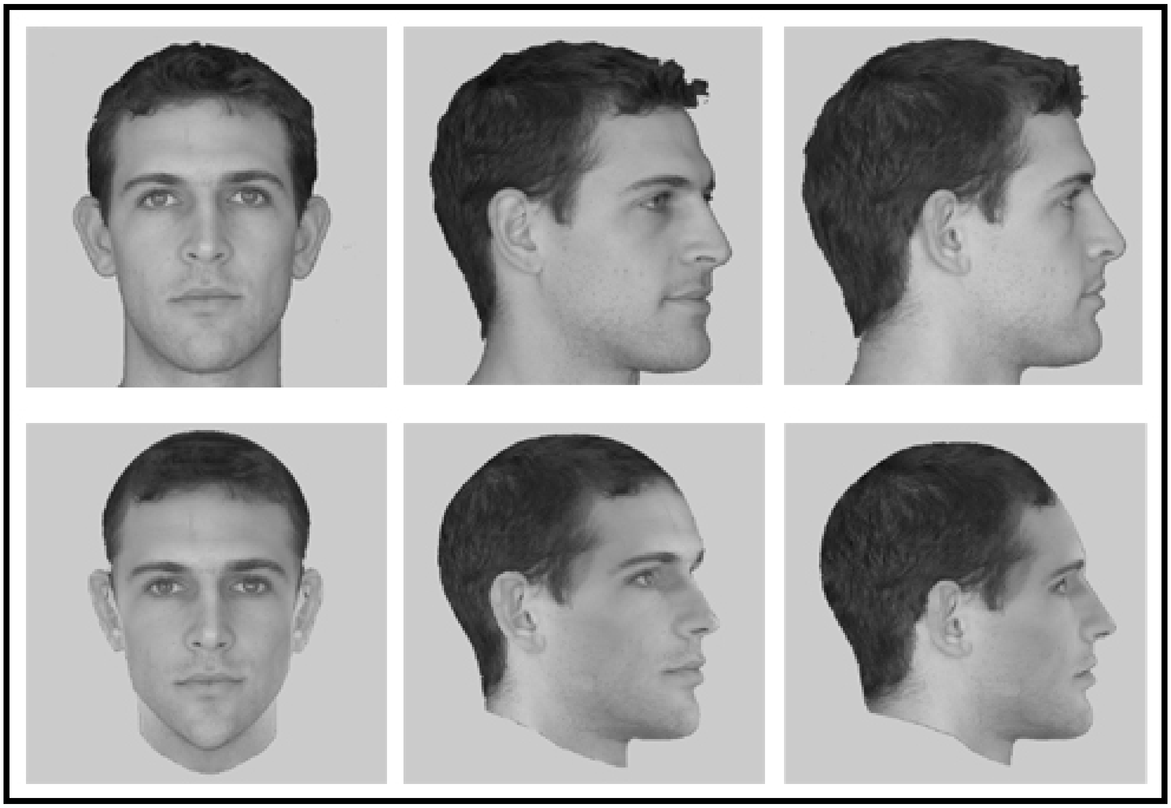

Figure 1: A sample stimulus from the study. The top row contains photographs; the bottom row contains screen-shots of the virtual model. Full views are on the left, ¾ views are in the middle, and profiles are on the right.

shows the six different conditions for a single face. The recognition test was forced choice, in that subjects were forced to pick one face or the other as a training face. During each trial, subjects saw two faces. All pairs were of the same experimental condition (e.g., virtual profiles were paired with one another), and all pairs contained one face that subjects had learned prior and one new distracter face. The pairs were presented side by side on the same computer screen. Subjects were instructed to choose the face on which they had been trained by hitting the ‘A’ key for training faces that appeared on the left side of the screen and the ‘L’ key for training faces that appeared on the right side. The faces remained on the screen until subjects chose, at which time a new pair of faces appeared.

For each training face, subjects were tested on all six lineup conditions depicted in Figure 1. Consequently there were 30 unique trials in one block. Subjects were tested on a total of 4 blocks (120 trials total). The order of trials within each block was random. Furthermore, appearance of test faces and distracter faces on screen-side (i.e., left vs. right) was random.

Results and Discussion. Figure 2 shows the means and confidence intervals for

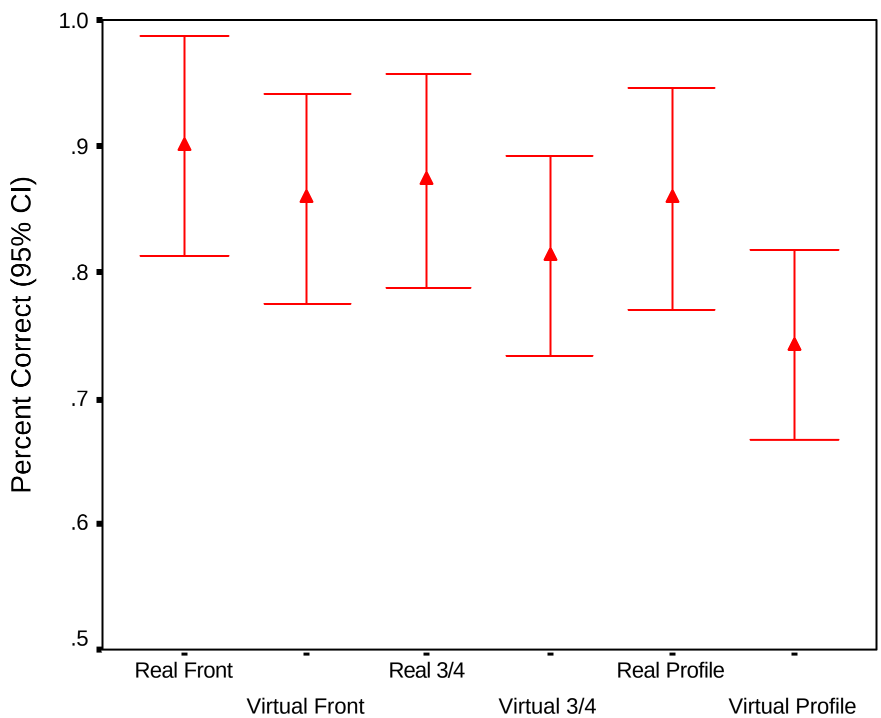

Figure 2: Mean percentage correct by condition in Experiment 1.

each condition. Both main effects were significant (face type: $F(1,24)=17.01$, $p<.001$, angle: $F(2,24)= 11.89$, $p<.001$). Subjects were slightly better overall with photographs than models (88% vs. 80%). Furthermore, there was a significant linear trend of angle, $F(1,25)=24.73$, $p<.001$, such that as the view moved farther from the frontal view performance was worse. This effect is similar to the recognition degradation seen using real human faces in experiments (Troje & Bueltoff, 1996). Performance was best (over 90%) in the photographic front view. The interaction between are two independent variables was also significant $F(2,24)=5.19$, $p>.05$, and was primarily driven by particularly poor recall in the virtual profile condition. This is consistent with previous research on photogrammetrically-generated virtual heads which demonstrates the largest structural (i.e., anthropometric) similarity from the profile angle (Bailenson, Beall, Blascovich, & Rex, 2003).

Future Directions

The results from the current study confirm that, while virtual heads promise to be quite an effective tool for person identification, photogrammetrically generated heads are not yet as effective as photographs in facilitating recognition of persons seen previously. However, in the current implementation, we used extremely inexpensive software and only dedicated approximately 45 minutes to the construction of each virtual head. Given more expenses in terms of time and money, the quality of photogrammetrically generated virtual heads should improve greatly.

In future work, we plan on using immersive virtual reality technology (i.e., Loomis, Blascovich, & Beall) to conduct lineups. In those studies, we plan on simulating crimes in the laboratory to naïve experimental subjects, and to then later test their recognition for the alleged suspect while the subjects (i.e., the identifiers) wear a stereoscopic head-mounted display that is capable of tracking their naturalistic movements (see www.worldviz.com for demonstrations of this technology). Wells (2002) points out a number of aspects in which the lineup process is currently problematic. We are optimistic that immersive virtual reality technology may provide a number of benefits to the lineup process.

References

Bailenson, J.N., Beall, A., C., Blascovich, J., & Rex, C. (2003). Examining Virtual Busts: Are Photogrammetrically-Generated Head Models Effective for Person Identification? Manuscript under review.

Bailenson, J.N., Wiggins, B., Blascovich, J., & Beall, A.C. (2003, accepted pending revision). Courtroom Applications of Virtual Environments, Immersive Virtual Environments, and Shared Virtual Environments. Journal of Law and Public Policy.

Blanz, V. & T. Vetter (1999). A Morphable Model for the Synthesis of 3D Faces. SIGGRAPH'99 Conference Proceedings, 187-194.

Busey, T. A. (1988). Physical and psychological representations of faces: Evidence from morphing. Psychological Science, 9, 476-483.

Brunelli, R. Falavigna, D., Poggio, T., & Stringa, L. (1995). Automatic Person Recognition by Using Acoustic and Geometric Features. Machine Vision and Applications, 8 p.317-325.

Decarlo, D., Metaxas, D., & Stone, M. (1998). An anthropometric face model using variational techniques. Proceedings of SIGGGRAPH '98, 67-74.

Liu, C. H., Collin, C. A., & Chaudhuri, A. (2000). Does face recognition rely on encoding of 3-D surface? Examining the role of shape-from-shading and shape-from-stereo. Perception, 29, 729-743.

Troje, NF and Bülthoff, HH (1996). Face recognition under varying pose: The role of texture and shape. Vision Research, 36, 1761-1771.

Wells, G. L. (2002). Eyewitness identification: Psychological aspects. The encyclopedia of crime and justice. New York: Macmillan.