Assessing and Improving the Identification of Computer Generated Portraits

OLIVIA HOLMES, Dartmouth College
MARTIN S. BANKS, University of California, Berkeley
and HANY FARID, Dartmouth College

Modern computer graphics are capable of generating highly photo-realistic images. While this can be considered a success for the computer graphics community, it has given rise to complex forensic and legal issues. A compelling example comes in the need to distinguish between computer-generated and photographic images as it pertains to the legality and prosecution of child pornography in the United States. We performed psychophysical experiments to determine the accuracy with which observers are capable of distinguishing computer-generated from photographic images. We find that observers have considerable difficulty performing this task, more difficulty than we observed five years ago when computer-generated imagery was not as photo-realistic. We also find that observers are more likely to report that an image is photographic than computer generated, and that resolution has surprisingly little effect on performance. Finally, we find that a small amount of training greatly improves accuracy.

Categories and Subject Descriptors: I.3.0 [**Computer Graphics**] General; H.1.2 [**Models and Principles**]: User/Machine Systems—*Human Information Processing*

General Terms: Human Factors, Legal Aspects

Additional Key Words and Phrases: computer graphics, photo-realistic, photo forensics

ACM Reference Format:

Holmes, O. and Banks, M.S. and Farid, H., C. 2015.Assessing and Improving the Identification of Computer Generated Portraits. ACM Trans. Appl. Percept. 2, 3, Article 1 (May 2010), 12 pages.
DOI = 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

1. INTRODUCTION

As 3-D rendering software and hardware have become increasingly more powerful, the computer-generated characters they create have become more photo realistic. There are many motivations for creating computer-generated characters that to the human viewer appear to be real: film making, video games, advertising, and much more. The goal of these applications is to create computer-generated characters that are indistinguishable from human characters. The achievement of this goal, however,

This work is supported by NSF BCS-1354029.
Author's address: Olivia Holmes olivia.b.holmes@gmail.com; Martin S. Banks martybanks@berkeley.edu; Hany Farid farid@cs.dartmouth.edu

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

can pose significant challenges. A clear and compelling example comes from the forensic and legal communities: The legality and prosecution of child pornography in the United States.

In the landmark 1982 case of *New York v. Ferber*, the US Supreme Court ruled that a New York state law banning child pornography was constitutional and not in violation of the First Amendment, and thus upholding that child pornography was illegal [nyf 1982]. In an attempt to respond to the growth of the on-line distribution of child pornography, the US Congress passed the Child Pornography Prevention Act (CPPA) in 1996 which made illegal "any visual depiction including any photograph, film, video, picture or computer-generated image that is, or appears to be, of a minor engaging in sexually explicit conduct" [cpp 1996]. In 2002, the CPPA was challenged in the US Supreme Court case of *Ashcroft v. Free Speech Coalition*. In this case, it was argued that the CPPA was overly broad and thus created an unintentional ban on protected speech. The Court agreed and found that portions of the CPPA were overly broad and infringed on the First Amendment. This new ruling classified computer-generated child pornography, the creation of which does not involve an actual child, as protected speech. As a result, a defense attorney need only claim that his client's images of child pornography are computer generated and thus protected material. The burden of proof then falls on the prosecutor to prove that the images are in fact real (photographic). The ability to distinguish between protected (computer-generated) and illegal (photographic) material, therefore, became essential to prosecuting child pornography crimes.

In 2003, seeking to remedy this unintended consequence that complicated the prosecution of child pornography, the US Congress passed the PROTECT Act which classified computer generated child pornography as "obscene" [pro 2003]. However, this law has not eliminated the challenge of responding to the so-called "virtual defense" (claiming that the material in question is computer generated) because juries are reluctant to send a defendant to jail on an obscenity charge for merely possessing computer-generated imagery when no real child was harmed in its creation. In contrast, possession of pornographic images in which real children are used is much easier to prosecute because the harm is direct and real. The ability to distinguish between computer-generated and photographic content, therefore, remains critical.

Over the past decade, some progress has been made in developing computational techniques to discriminate computer-generated from real characters. Most of these techniques exploit regularities in some low- to mid-level statistical features extracted from computer-generated and natural images. The first of such approaches to this problem used statistical features extracted from the wavelet domain [Lyu and Farid 2005; Wang and Moulin 2006]. Related techniques leveraged sensor noise [Dehnie et al. 2006; Khanna et al. 2008], demosaicing artifacts [Dirik et al. 2007], chromatic aberrations [Gallagher and Chen 2008], geometric- and physics-based image features [Ng et al. 2005], and color compatibility [Lalonde and Efros 2007; Wen et al. 2007]. More recently, non-statistically based techniques have been proposed that exploit facial asymmetries [Dang-Nguyen et al. 2012a], repetitive patterns of facial expressions [Dang-Nguyen et al. 2012b], and a measure of facial blood flow [Conotter et al. 2014]. These techniques, however, are yet to be widely adopted by the courts and it is far more typical for the courts to rely on jurors to assess the origin of images.

It is therefore important to ask: How well can the average juror distinguish a computer-generated from a photographic image? Previous work found that human observers can reliably distinguish computer-generated from photographic portraits of people [Farid and Bravo 2007; 2012; Fan et al. 2012]. In this study, we seek to determine if advances in computer graphics over the past several years have affected an observer's ability to perform this task. We also describe a simple way in which observer accuracy at this task can be significantly improved. In addition to the legal implications, the results of this study should be of interest to the computer graphics community as a measure of their success in creating photo-realistic imagery, as well as those interested in the nature of animacy and what makes a person in a photograph seem alive [Looser and Wheatley 2010].

2. METHODS

We first describe the collection and pre-processing of matching computer-generated and photographic images, after which we will describe the details of the experimental procedure.

2.1 Images

We downloaded 30 high-quality computer-generated images from the computer graphics websites: `www.cgsociety.org`, `www.3dtotal.com`, `www.cgarena.com`, and `www.romans3d.ru`. We selected these websites because they contained a large number of high-quality computer-generated images. The content and context of these websites virtually guaranteed that these images were computer generated. We received written confirmation from the website editors that the posted images were solely computer generated (i.e., did not contain any photographic components). Additionally, none of these 30 images contained metadata consistent with a photograph recorded by a digital camera. Each image contained a human face posed facing forward, a resolution of at least 800 pixels (defined as the minimum of its width and height), and a render date between 2013 and 2015 (with the majority created since 2014). The 30 computer-generated images are composed of 15 male and 15 female faces. Because we planned to ask observers to identify the gender of the person in the image, we chose images for which this was easily identifiable.

For each of the 30 computer-generated images, we found a matching photographic image in terms of age, gender, race, pose, and accessories. The majority of these photographic images were downloaded from `www.flickr.com`. The content and context of these websites virtually guaranteed that these images were photographic. Additionally, 11 of these 30 images contained metadata consistent with a photo recorded by a digital camera. With respect to the remaining 19 images, it is not unusual for digital images to be stripped of their metadata prior to being uploaded to photo-sharing sites, so we think it is extremely likely that they too are photographs. These matched computer-generated and photographic images are shown in Figures 1– 3.

All computer-generated and photographic images were matched for brightness and contrast (in luminance/chrominance space) to ensure that observers could not classify images based on systematic differences in a low-level image statistic (as they could if, for example, the computer-generated images generally had a higher contrast than their photographic counterparts). Each computer-generated and photographic RGB image was color adjusted to match the mean and variance of each luminance and chrominance channel. Denote a computer-generated image as $f_g(x, y, c)$ and a photographic image as $f_p(x, y, c)$, where c corresponds to the luminance (Y), chrominance (Cb), or chrominance (Cr) channel:

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix} + \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{1}$$

The mean of each YCbCr channel was set to $\mu(c)$:

$$\hat{f}_g(x, y, c) = f_g(x, y, c) - \mu_g(c) + \mu(c) \tag{2}$$

$$\hat{f}_p(x, y, c) = f_p(x, y, c) - \mu_p(c) + \mu(c) \tag{3}$$

where $\mu_g(c)$ and $\mu_p(c)$ are the means of the c^{th} channel of the individual computer-generated and photographic images, and $\mu(c)$ is the average mean across all 60 images ($\vec{\mu} = [106, 118, 141]$). The variance

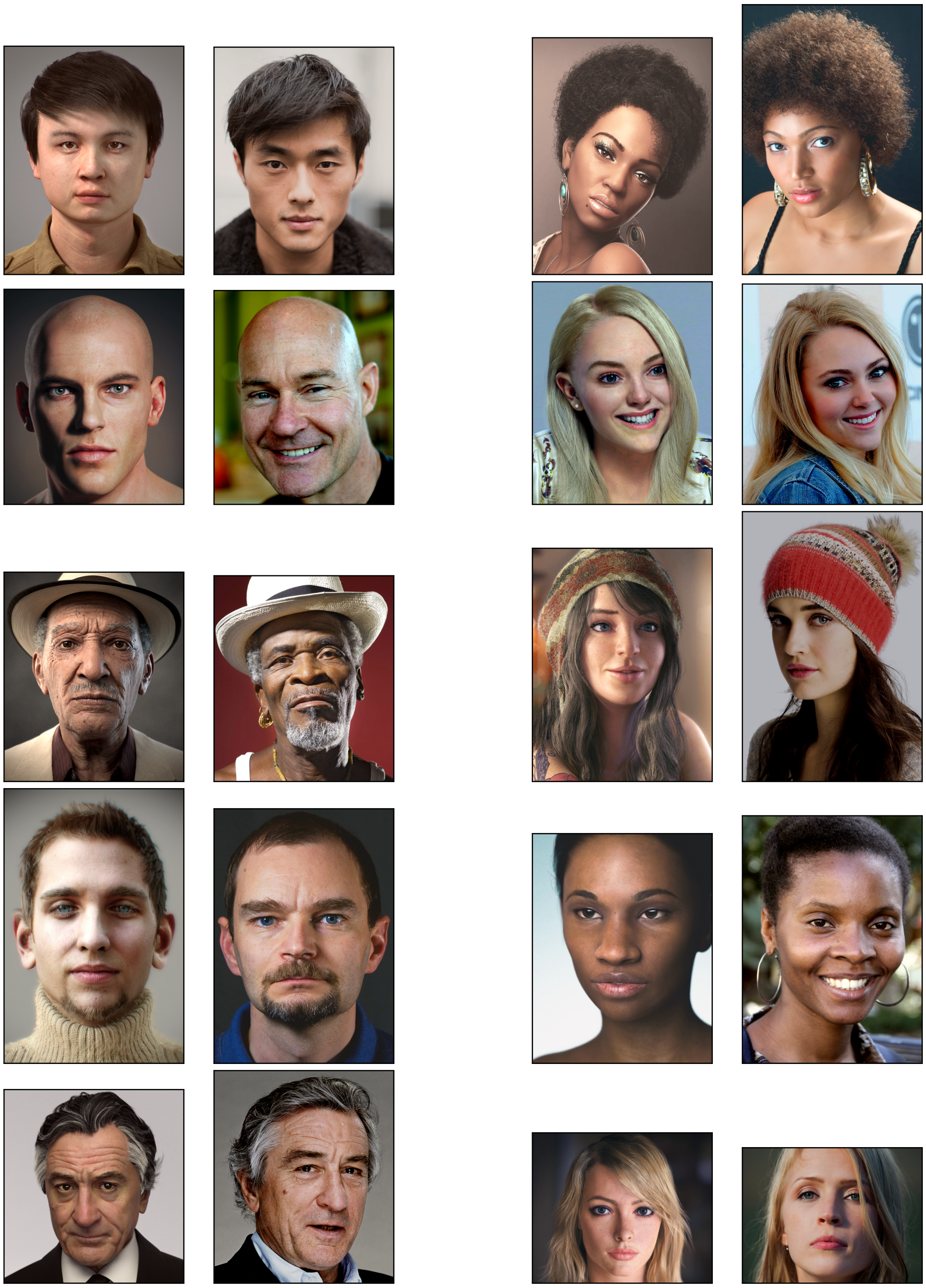


Fig. 1. Computer-generated images (first and third columns) paired with their photographic matches (second and fourth columns). See Figure 2 and 3.

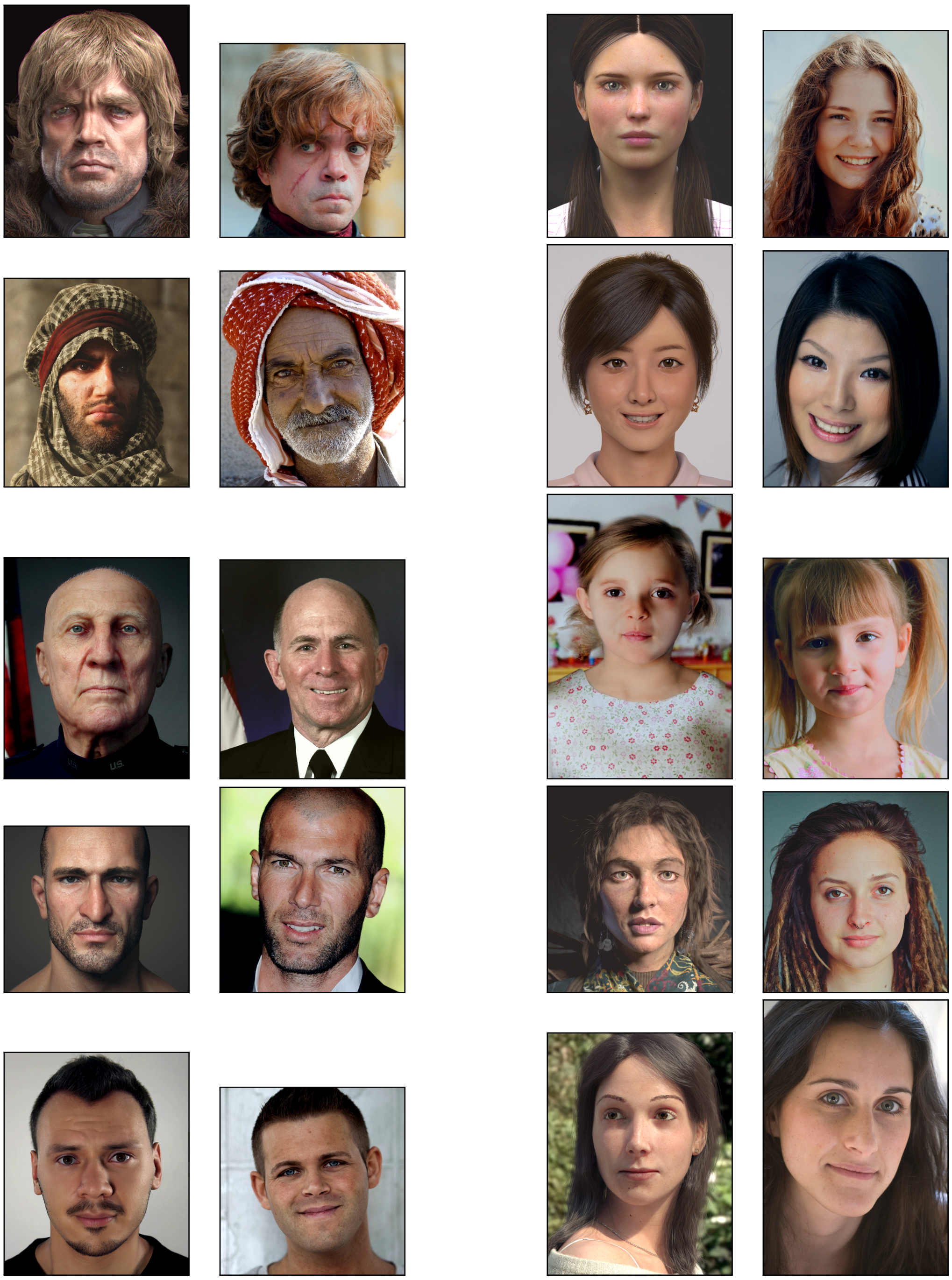


Fig. 2. Computer-generated images (first and third columns) paired with their photographic matches (second and fourth columns). See Figure 1 and 3.

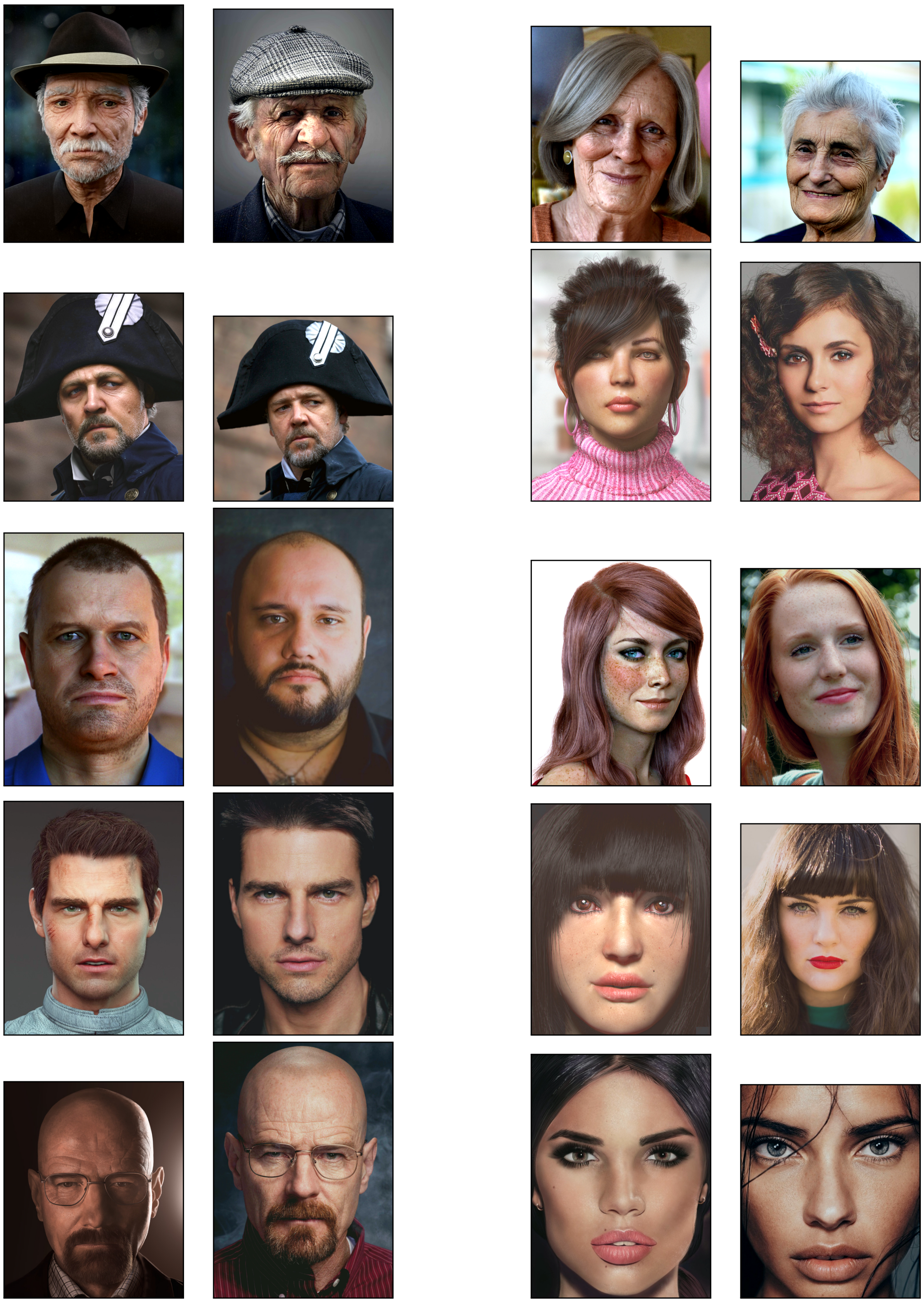


Fig. 3. Computer-generated images (first and third columns) paired with their photographic matches (second and fourth columns). See Figure 1 and 2.

of each channel was then set to $\sigma(c)$:

$$\hat{f}_g(x,y,c) \;=\; \sqrt{\frac{\sigma(c)}{\sigma_g}}(\hat{f}_g(x,y,c)-\mu(c))+\mu(c) \tag{4}$$

$$\hat{f}_p(x,y,c) \;=\; \sqrt{\frac{\sigma(c)}{\sigma_p}}(\hat{f}_p(x,y,c)-\mu(c))+\mu(c) \tag{5}$$

where σ_g and σ_p are the variances of the c^{th} channel of the individual computer-generated and photographic image, and $\vec{\sigma}$ is the average variance across all 60 images ($\vec{\sigma} = [2299, 54, 92]$). After color adjusting in the luminance/chrominance space, the images were converted back to their original RGB color space:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix} + \begin{pmatrix} 1.000 & 0.000 & 1.400 \\ 1.000 & -0.343 & -0.711 \\ 1.000 & 1.765 & 0.000 \end{pmatrix} \begin{pmatrix} Y \\ Cb-128 \\ Cr-128 \end{pmatrix} \tag{6}$$

Although we found no systematic differences in the original brightness and contrast of the computer-generated and photographic images, this image matching ensured that observers were not distracted by this low-level cue.

2.2 Experimental Procedure

We recruited participants from Amazon's Mechanical Turk on-line work force. After a Mechanical Turk user opted to participate in our experiment, they were given a brief summary of the task and were asked to consent to the terms of the experiment.

During the experiment, an observer viewed a total of 60 images, 10 at each of six resolutions (100, 200, 300, 400, 500, or 600 pixels, measured as the largest dimension). An observer never viewed the same image, regardless of resolution, more than once. After each image was presented, the observer was asked to make a judgment as to whether the image was computer generated (CG) or photographic and whether the person in the image was female or male. To ensure that observers did not rush their judgment, the experimental program did not allow an observer to make their selection for 3 seconds after the image onset. After this delay, the observer could click a button to indicate their choice: “male/CG”, “male/photographic”, “female/CG”, or “female/photographic”.

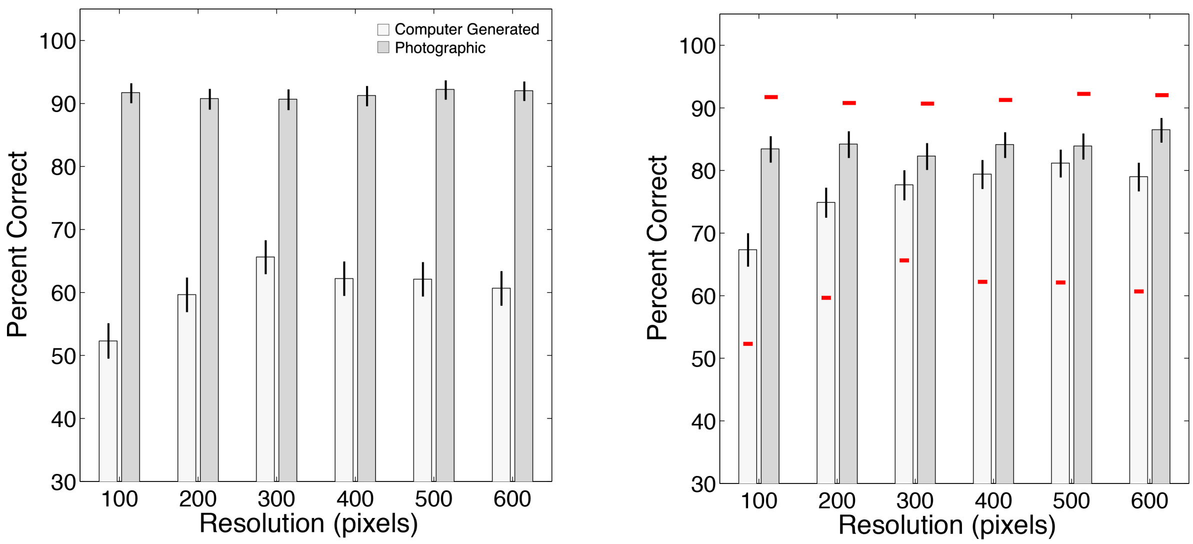
The order in which the 60 images were presented to the observer was randomized. Each participant was paid \$0.50 for their time. No feedback was given to participants indicating how well they had performed. The observer's ability to correctly identify the person's gender in each image was used as a means of discarding the results from observers who responded randomly. A threshold of 95% accuracy on this gender task was set prior to data collection.

3. RESULTS

We performed two experiments. In the first, observers classified an image as computer generated or photographic. Observer performance was assessed as a function of image resolution. The second experiment was similar except this task was preceded by a short training session in which observers were shown representative examples of computer-generated and photographic images. Except where indicated, the procedures in these experiments were identical.

3.1 Experiment 1: No Training

We collected responses from 250 Mechanical Turk participants. Only one participant out of 250 was excluded because their accuracy in determining the gender of the person in the image was below our



(a) no training (b) training

Fig. 4. Observer accuracy as a function of resolution for Experiment 1 (no training) and Experiment 2 (training). The light gray bars indicate performance on computer-generated images and the dark gray bars indicate performance on photographic images. Error bars shown in black represent 95% confidence intervals. The small horizontal red bars in panel (b) denote the performance in the no-training condition shown in panel (a). Chance performance is 50%. See also Table I.

threshold. Shown in Figure 4(a) is the average observer accuracy in correctly identifying computer-generated and photographic images as a function of image resolution. Percent correct is probably not the most informative measure of performance because observers may be biased to report photographic more frequently than computer generated. For example, if an observer had no ability to make the discrimination, but always responded photographic, their percent correct with photographic images would be 100%. In order to better handle such potential biases, we used tools from signal detection theory [Green and Sweats 1966]. Shown in Table I(a) are these results expressed as d' and β. d' represents observer sensitivity (independent of bias) and β represents observer bias.[1] Higher values of d' correspond to greater observer sensitivity. A value of $d' = 0$ means that the observer has no information to make reliable identifications no matter what bias he or she might have. A value of $\beta = 1.0$ would indicate no bias, a value of $\beta > 1$ indicates that observers are biased to classifying an image as photographic, and $\beta < 1$ indicates that observers are biased to classifying an image as computer generated.

Observers have a clear bias for classifying an image as photographic. Across all six resolutions, the average β is 2.46. Resolution has only a small impact on performance from d' of 1.68 (at the maximum resolution of 600) to d' of 1.45 (at the minimum resolution of 100).

As described in [Farid and Bravo 2012], observer accuracy for images rendered prior to 2010 yielded considerably higher accuracy $d' = 2.46$, but similar bias $\beta = 2.37$. Despite the overall reduction in accuracy in the intervening five years, 12.4% of our participants had a d' greater than 2.50. They also

[1]Denote the hit-rate H as the accuracy of correctly identifying a computer-generated image and the false-alarm rate F as 1 minus the accuracy of correctly identifying a photographic image. Observer sensitivity d' is $z(H) - z(F)$, where $z(\cdot)$ corresponds to the z-transform. Observer bias β is $g(z(H))/g(z(F))$, where $g(\cdot)$ is a zero-mean, unit-variance normal distribution.

Table I. d' and β as a function of resolution for Experiment 1 (no training) and Experiment 2 (training). See also Figure 4 and 6.

	(a) no training		(b) training	
resolution	d'	β	d'	β
100	1.45	2.61	1.42	1.45
200	1.57	2.34	1.68	1.32
300	1.73	2.21	1.69	1.15
400	1.67	2.40	1.82	1.18
500	1.73	2.62	1.88	1.11
600	1.68	2.60	1.91	1.33

had a larger bias on average ($\beta = 3.13$). Not surprisingly, advances in photo-realistic rendering have made it more difficult for observers to distinguish the computer generated from the photographic.

3.2 Experiment 2: Training

In this second experiment we attempted to eliminate or minimize the bias observed in Experiment 1. To do so, a new set of observers was shown a new set of 10 computer-generated and 10 matching photographic images (Figure 5) each of which was labeled as either computer generated or photographic. Observers passively viewed each image for at least 3 seconds before pressing a key to view the next image. After viewing these 20 training images, observers viewed the same 60 images used in Experiment 1.

We collected responses from 250 Mechanical Turk participants (each of whom did not participate in Experiment 1). Only three participants out of 250 were excluded because their accuracy in determining the gender of the person in the image was below our 95% threshold. Shown in Figure 4(b) is the average observer accuracy in correctly identifying computer-generated and photographic images as a function of image resolution, and shown in Table I(b) are these results expressed as d' and β. The per-image accuracies are shown in Figure 6.

The results show that a small amount of training was surprisingly effective at significantly reducing the bias found in Experiment 1. Across all resolutions, the average β fell from 2.46 to 1.51. The average d' increased slightly from 1.64 to 1.73, and at the highest resolution of 600, it increased from 1.68 to 1.91.

Similar to Experiment 1, our top-performing observers have a considerably higher sensitivity: 21.9% had a d' greater than 2.50, with an average bias slightly higher than that of the entire group (β of 1.72). The number of high-performing observers was nearly double as compared to Experiment 1.

4. DISCUSSION

The past five years has seen tremendous progress in the creation of highly photo-realistic imagery. It is therefore perhaps not surprising that the average observer now struggles to distinguish a photographic from a modern-day computer-generated portrait of a person. Observer accuracy in recognizing modern-day computer-generated images is significantly worse than it was five years ago [Farid and Bravo 2012]. However, a significant bias to classify an image as photographic still persists amongst human observers after five years time, which can be quite problematic in a legal setting where this distinction can dramatically change the nature of a criminal charge.

We found, however, that this bias can be nearly eliminated with a small amount of training in which observers are shown representative examples of computer-generated and photographic images. After training, average observer accuracy hovers around 80% for both computer-generated and photographic images. It seems likely that this accuracy is a lower bound on human performance: Our observers were

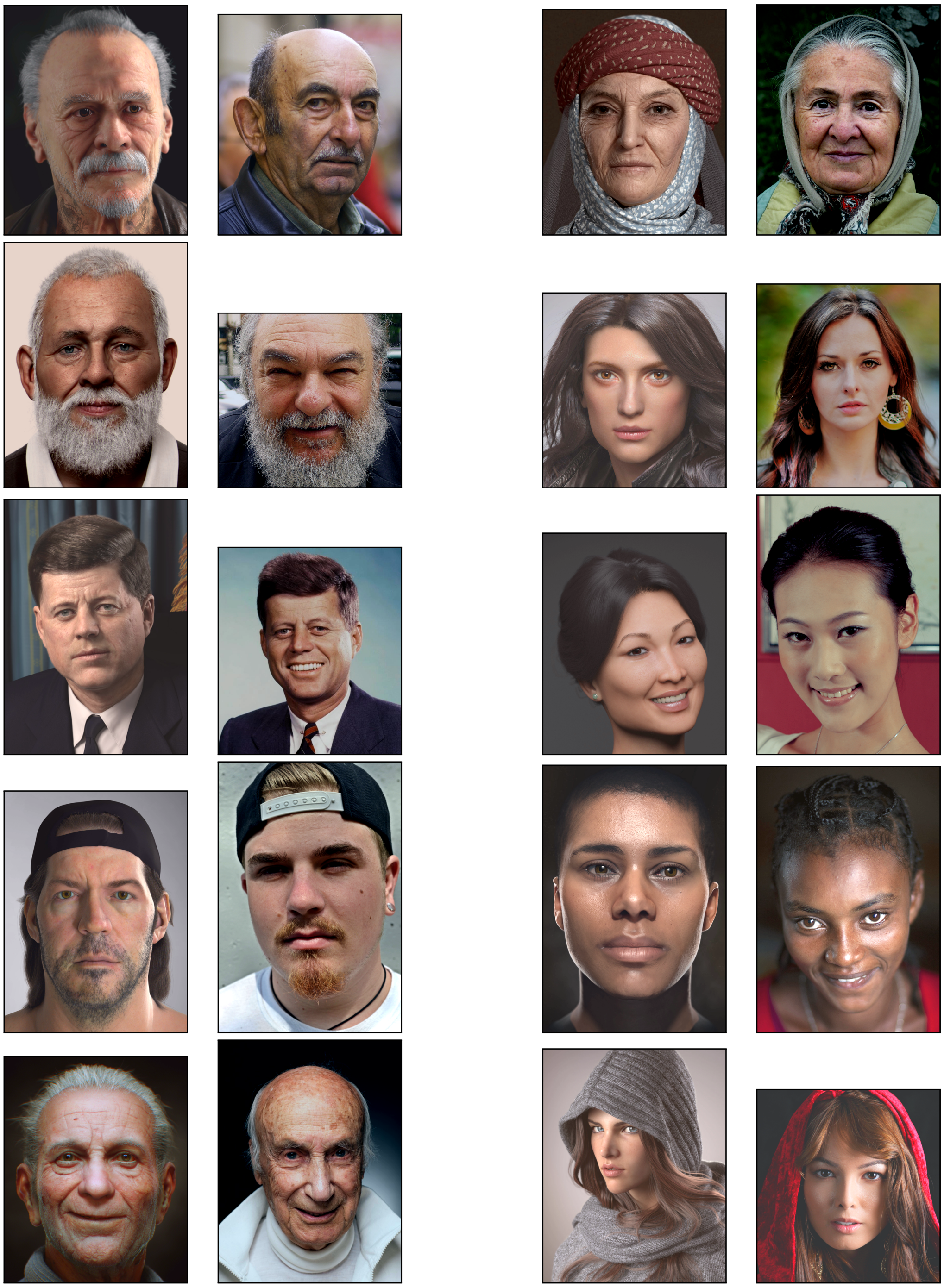


Fig. 5. Training computer-generated images (first and third columns) paired with their photographic matches (second and fourth columns).

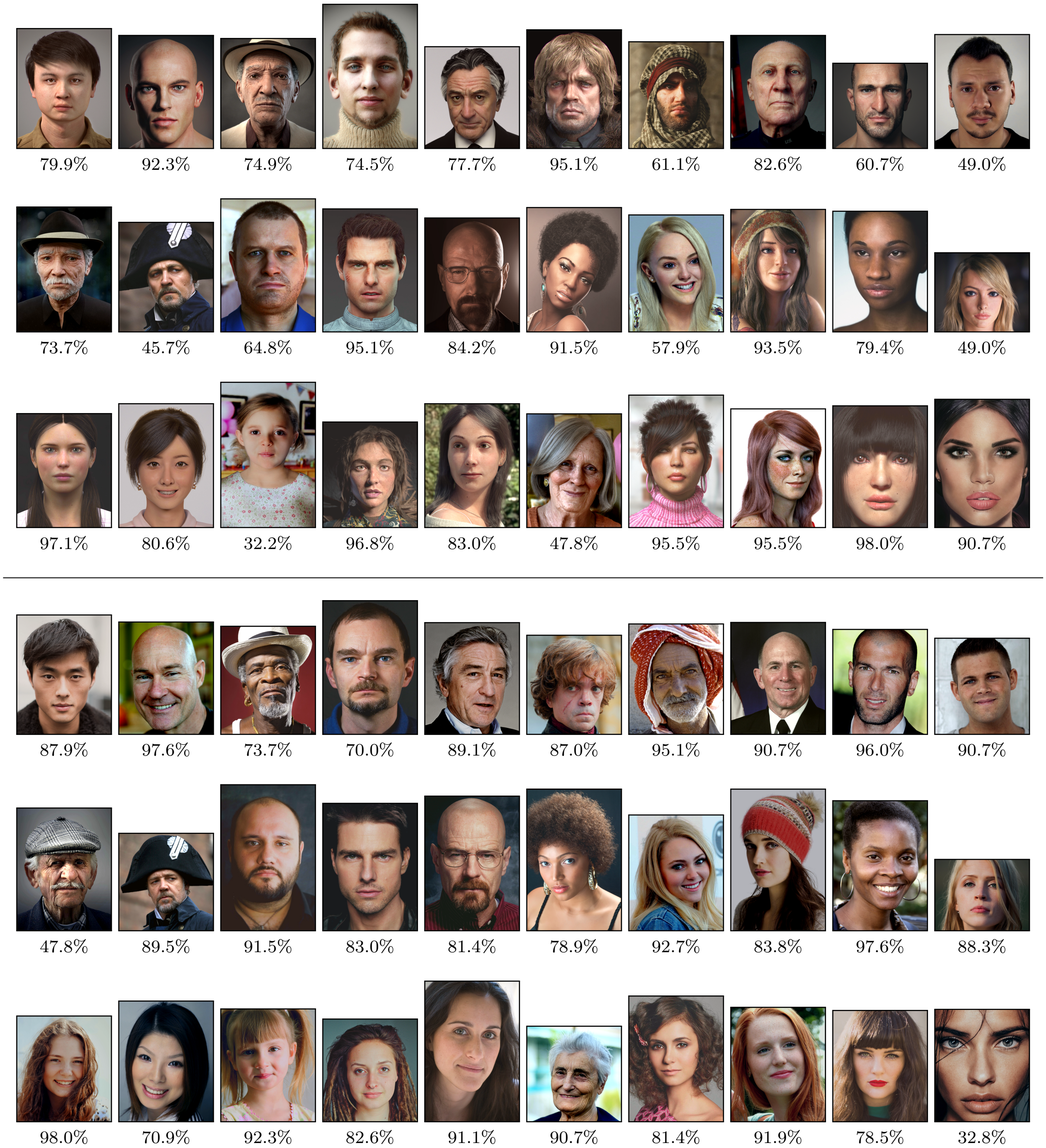


Fig. 6. Per-image classification accuracy for Experiment 2 (top: computer-generated images; bottom: photographic images).

given no incentive to perform well (their reward was independent of performance), virtually all of their decisions were made within 5 seconds, and, compared with the types of images encountered in forensic settings, our images were relatively impoverished, containing only a single person depicted from the neck up. A full body figure interacting with the environment or other people is far more difficult to render photo-realistically, and presumably would be easier for observers to identify.

We expect that as computer-graphics technology continues to advance, observers will find it increasingly difficult to distinguish computer-generated from photographic images. While this can be considered a success for the computer-graphics community, it will no doubt lead to complications for the legal and forensic communities. We expect that human observers will be able to continue to perform this task for a few years to come, but eventually we will have to refine existing techniques and develop new computational methods that can detect fine-grained image details that may not be identifiable by the human visual system.

REFERENCES

1982. New York v. Ferber.

1996. Child Pornography Prevention Act (CPPA).

2003. Prosecutorial Remedies and Other Tools to end the Exploitation of Children Today (PROTECT) Act.

CONOTTER, V., BODNARI, E., GIULIA, B., AND FARID, H. 2014. Physiologically-based detection of computer generated faces in video. In *IEEE International Conference on Image Processing*.

DANG-NGUYEN, D.-T., BOATO, G., AND DE NATALE, F. G. B. 2012a. Discrimination between computer generated and natural human faces based on asymmetry information. In *IEEE European Signal Processing*. 1234–1238.

DANG-NGUYEN, D.-T., BOATO, G., AND DE NATALE, F. G. B. 2012b. Identify computer generated characters by analysing facial expressions variation. In *IEEE Workshop on Information Forensics and Security*. 252–257.

DEHNIE, S., SENCAR, H., AND MEMON, S. 2006. Digital image forensics for identifying computer generated and digital camera images. In *IEEE International Conference on Image Processing*. 2313–2316.

DIRIK, A. E., BAYRAM, S., SENCAR, H. T., AND MEMON, N. 2007. New features to identify computer generated images. In *IEEE International Conference on Image Processing*. Vol. 4.

FAN, S., NG, T.-T., HERBERG, J. S., KOENIG, B. L., AND XIN, S. 2012. Real or fake?: Human judgments about photographs and computer-generated images of faces. In *SIGGRAPH Asia 2012 Technical Briefs*. 17:1–17:4.

FARID, H. AND BRAVO, M. J. 2007. Photorealistic rendering: How realistic is it? *Journal of Vision 7,* 9.

FARID, H. AND BRAVO, M. J. 2012. Perceptual discrimination of computer generated and photographic faces. *Digital Investigation 8*, 226–235.

GALLAGHER, A. C. AND CHEN, T. 2008. Image authentication by detecting traces of demosaicing. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 1–8.

GREEN, D. AND SWEATS, J. 1966. *Signal Detection Theory and Psychophysics*. Peninsula Pub.

KHANNA, N., CHIU, G. T. C., ALLEBACH, J. P., AND DELP, E. J. 2008. Forensic techniques for classifying scanner, computer generated and digital camera images. In *IEEE International Conference on Acoustics, Speech and Signal Processing*. 1653–1656.

LALONDE, J. F. AND EFROS, A. A. 2007. Using color compatibility for assessing image realism. In *IEEE International Conference on Computer Vision*. 1–8.

LOOSER, C. AND WHEATLEY, T. 2010. The tipping point of animacy: How, when, and where we perceive life in a face. *Psychological Science 21*, 1854–1862.

LYU, S. AND FARID, H. 2005. How realistic is photorealistic? *IEEE Transactions on Signal Processing 53,* 2, 845–850.

NG, T.-T., CHANG, S.-F., HSU, J., XIE, L., AND TSUI, M.-P. 2005. Physics-motivated features for distinguishing photographic images and computer graphics. In *ACM International Conference on Multimedia*. 239–248.

WANG, Y. AND MOULIN, P. 2006. On discrimination between photorealistic and photographic images. In *IEEE International Conference on Acoustics, Speech and Signal Processing*. Vol. 2.

WEN, C., SHI YUN, Q., AND GUORONG, X. 2007. Identifying computer graphics using HSV color model. In *IEEE International Conference on Multimedia and Expo*.