A Case Study of Automated Face Recognition: The Boston Marathon Bombings Suspects

Joshua C. Klontz, *Noblis*

Anil K. Jain, *Michigan State University*

Commercial automated face recognition technology shows promise in a challenging identification environment.

Researchers have made tremendous advances in automated face recognition over the past two decades. State-of-the-art commercial systems can match face images in mug shot, passport, and driver's license databases with nearly perfect accuracy. Consequently, the Federal Bureau of Investigation and many state and local law enforcement agencies across the US routinely use face recognition technology to identify criminal suspects.

However, factors such as ambient illumination and differences in pose, expression, and age between query and reference images can confound face recognition algorithms. As Figure 1 shows, when the query subject is in an unconstrained imaging environment, the true accept rate can fall from 99 percent to below 60 percent (J. Klontz et al., "Open Source Biometric Recognition," to appear in *Proc. 6th Int'l Conf. Biometrics: Theory, Applications, and Systems* [BTAS 13], IEEE, 2013). These challenges make it more difficult to leverage automated face recognition for forensic applications using low-resolution query images obtained from CCTV or mobile devices.

In early August 2011, four days of riots across England stemming from the police shooting of Mark Duggan, a resident of Tottenham, North London, led to widespread looting, arson, and violent clashes with authorities. Despite substantial CCTV coverage, law enforcement was unable to rely on face recognition technology to identify rioters and instead appealed to the public for help by broadcasting video footage on local news stations.

US authorities faced a similar dilemma in the wake of the Boston Marathon bombings on 15 April 2013. Despite extensive video footage and images from local surveillance systems and onlookers' cameras and smartphones, local, state, and federal law enforcement agencies were unable to identify the two suspects, Dzhokhar and Tamerlan Tsarnaev, using only face recognition technology. They ultimately had to turn to crowdsourcing to identify the brothers, which occasioned a massive manhunt, the shutdown of Watertown, Massachusetts—a city of more than 30,000 people—and a deadly firefight with police as the suspects tried to evade arrest.

CASE STUDY

While face recognition can be a useful tool, helping authorities to narrow leads and confirm a criminal suspect's identity, the media

0018-9162/13/$31.00 © 2013 IEEE Published by the IEEE Computer Society

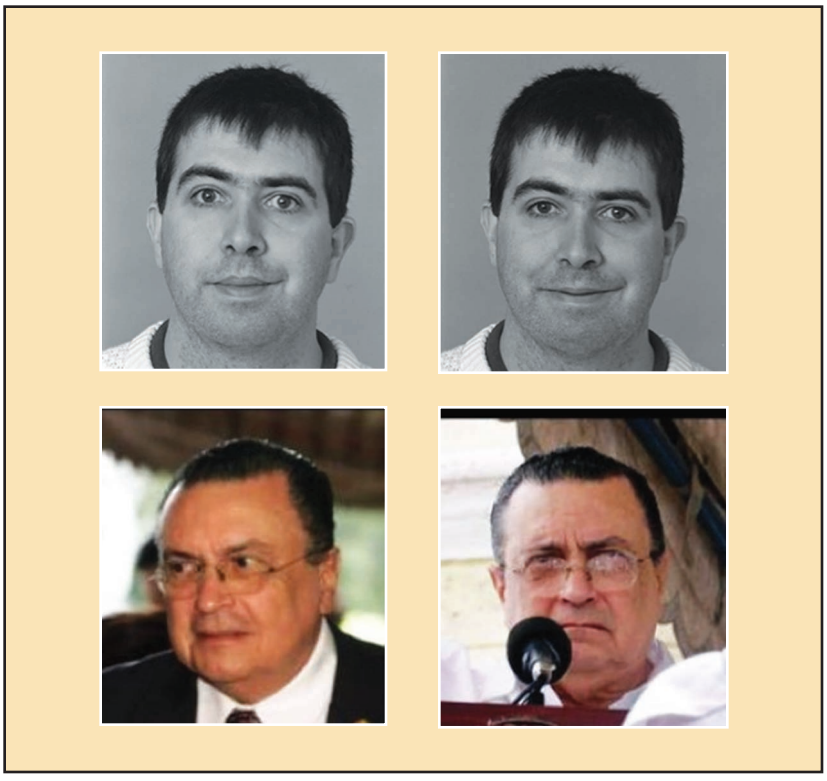

Figure 1. Automated face recognition systems can match images of cooperative subjects in controlled environments (top) with a 99 percent true accept rate, but accuracy falls to below a 60 percent true accept rate when subjects in the query image are in unconstrained environments (bottom). In both cases, the false accept rate is 0.1 percent. (Photo sources: top row—P.J. Phillips et al. "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, 2000, pp. 1090-1104; bottom row—G.B. Huang et al., *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*, tech. report 07-49, Computer Science Dept., Univ. of Massachusetts Amherst, Oct. 2007.)

tended to highlight the technology's limitations in the Boston Marathon bombings investigation. Consequently, we were motivated to reexamine the efficacy of unconstrained face recognition using this incident as a case study.

We chose two automated face recognition systems known for their superior unconstrained matching performance based on tests conducted by the US National Institute of Standards and Technology: NEC's NeoFace v3.1 (www.necam.com/Biometrics/doc.cfm?t=FaceRecognition) and Google-owned Pittsburgh Pattern Recognition (PittPatt) v5.2.2.

To conduct the experiment, we added six reference images of the Tsarnaevs taken from press releases and news articles following their identification, shown in Figure 2, to a background database of 1 million mug shots provided by the Pinellas County (Florida) Sheriff's Office, which has been using face recognition technology for over a decade. Against this database, we searched for matches of the five face images of the brothers extracted by the FBI from surveillance, smartphone, or point-and-shoot camera footage before their identification, shown in Figure 3.

NeoFace outperformed PittPatt for most of the searches, though both algorithms were able to extract salient features from the relatively noisy query images. Probes for the younger brother, Dzhokhar Tsarnaev, exhibited notably better retrieval rates than those for Tamerlan Tsarnaev, whose face in the query images is occluded by sunglasses. The middle query image of Dzhokhar Tsarnaev in Figure 3b, which appears to be a modified version of the first query image, generally resulted in lower matching accuracy than its unmodified counterpart, likely due to the alteration.

Both face recognition systems had the most success matching the query images to the middle reference images in Figures 2a and 2b. Pose consistency between compared faces was likely the crucial factor: in these reference images, the faces are turned slightly to the left. In particular, NeoFace returned the middle reference image of Dzhokhar Tsarnaev in Figure 2b as a rank-one hit for the third query image in Figure 3b.

We then compared each probe image to reference images of subjects in the database with physical characteristics that matched descriptions of the bombings suspects: white, male, and 20 to 30 years old in the case of Tamerlan Tsarnaev, and white, male, and 15 to 25 years old in the case of Dzhokhar Tsarnaev. Demographic filtering reduced the size of the reference gallery from 1 million to 174,718 and 131,462 images, respectively. This substantially improved retrieval rankings, generally proportional to the reduction in gallery size.

Finally, we ran a fused search that summed up the match scores of searches using different query images of the same suspect before ranking the gallery images. In general, fusion improved retrieval rates for gallery images ranked similarly by each probe but lowered retrieval rates for gallery images ranked differently across the fused probes.

Figure 2. Reference face images of the Boston Marathon bombings suspects from press releases and news articles following their identification: (a) Tamerlan Tsarnaev and (b) Dzhokhar Tsarnaev.

Figure 3. Query face images of the Boston Marathon bombings suspects from media released by the FBI before their identification. (a) Tamerlan Tsarnaev and (b) Dzhokhar Tsarnaev.

LOOKING AHEAD

While the Boston Marathon bombings case offers only a small number of published query face images to assess the effectiveness of automated face recognition technology, our study yielded some useful insights.

Though not yet sufficiently accurate for turnkey deployment in law enforcement applications, the technology has made considerable

progress in identifying faces in unconstrained environments. NeoFace's rank-one hit of the middle reference image of Dzhokhar Tsarnaev in Figure 2b for the third query image in Figure 3b illustrates this potential. However, the hit was against a graduation photo with a similar pose, acquired after the suspect had been publicly identified, and not the kind of front-facing portrait found in government and law enforcement databases.

If the digital face evidence from the Boston Marathon bombings is representative of other scenarios, it might be time to rethink some of the classic confounding factors in automated facial recognition. In particular, illumination, expression, and aging appear to be less relevant as sources of error than image resolution and occlusion in our study. However, differences in head pose continue to challenge state-of-the-art commercial systems.

Re-identification—the process of discovering other instances of a subject within a dataset—and representation construction from multiple viewpoints are two symbiotic areas in need of further research. For example, in our study it likely would have proven instrumental to search for other instances of the "suspect in the white hat" across all the collected digital evidence. Ideally, information provided from multiple viewpoints could then be fused together to construct a single face representation of the suspect prior to searching against a large face database of known individuals. Recent work (L. Best-Rowden et al., "Video-to-Video Face Matching: Establishing a Baseline for Unconstrained Face Recognition," to appear in *Proc. 6th Int'l Conf. Biometrics: Theory, Applications, and Systems* [BTAS 13], IEEE, 2013) provides a preliminary study in this direction.

Joshua C. Klontz *is a computer vision researcher at Noblis, a nonprofit scientific research corporation based in Falls Church, Virginia. Contact him at Joshua.Klontz@noblis.org.*

Anil K. Jain *is a University Distinguished Professor in the Department of Computer Science and Engineering at Michigan State University. Contact him at jain@cse.msu.edu.*

Editor: Karl Ricanek Jr., director of the Face Aging Group at the University of North Carolina Wilmington; ricanekk@uncw.edu