A Case Study on Unconstrained Facial Recognition Using the Boston Marathon Bombings Suspects

Joshua C. Klontz
Michigan State University
East Lansing, MI, U.S.A
klontzjo@msu.edu

Anil K. Jain
Michigan State University
East Lansing, MI, U.S.A
jain@cse.msu.edu

Technical Report MSU-CSE-13-4
May 22, 2013

Abstract

The investigation surrounding the Boston Marathon bombings was a missed opportunity for automated facial recognition to assist law enforcement in identifying suspects. We simulate the identification scenario presented by the investigation using three state-of-the-art commercial face recognition systems, and evaluate the maturity of face recognition technology in matching low quality face images of uncooperative subjects. Our experimental results show one instance where a commercial face matcher returns a rank-one hit for suspect Dzhokhar Tsarnaev against a one million mugshot background database. Though issues surrounding pose, occlusion, and resolution continue to confound matchers, there have been significant advances made in face recognition technology to assist law enforcement agencies in their investigations.

1. Introduction

On April 15, 2013 at 2:49 p.m. EDT, two bombs exploded near the finish line of the Boston Marathon, killing 3 people and injuring 264 others [16]. The race was abruptly halted and police cornered off a 12-block crime scene surrounding the location of the blasts [17]. The Federal Bureau of Investigation (FBI) took the lead, and initial forensic evidence indicated the explosive device was a pressure cooker packed with fragments of BBs and nails, possibly concealed in a dark-colored nylon backpack [2].

Shortly after the bombing, more than 1,000 law enforcement officers across many agencies began canvassing sources, reviewing government and public databases, and conducting interviews with eyewitnesses [2]. Businesses were asked to review and preserve surveillance video and police received a "huge amount of video evidence" from the public [25].

After reviewing "photo, video, and other evidence" [3], the FBI released images and videos of the two suspects shown in Figure 1. In addition to seeking identification help, the release of the images and videos was also in part to limit the damage being done to people wrongly targeted as suspects by news and social media. Shortly after the release, the two suspects were identified as brothers, Tamerlan Tsarnaev and Dzhokhar Tsarnaev, by their aunt who made a call to the FBI tip line [19].

It is believed that the release of their photographs provoked the brothers into further violence, fatally shooting an MIT campus officer and carjacking a Mercedes SUV [19]. These events intensified the manhunt for the brothers that ultimately ended in a violent confrontation with police officers where Tamerlan Tsarnaev was killed and Dzhokhar Tsarnaev was wounded and later captured.

The investigation of the Boston Marathon bombings, outlined in Figure 2, has been widely viewed by the media as a failure for automated facial recognition [5,8]. The technology came up empty even though both Tsarnaevs' photos exist in official government databases: Dzhokhar had a Massachusetts driver's license; the two brothers had legally immigrated to the United States; and Tamerlan had been the subject of an FBI investigation [19].

This paper presents a case study in unconstrained facial recognition, using public domain images of the two suspects in the Boston Marathon bombings. Suspects' photographs are matched against a background set of mugshots with three state-of-the-art commercial face recognition systems. Results are used to gauge the maturity of available technology in unconstrained facial recognition scenarios.[1]

[1] In contrast to conventional face recognition, unconstrained recognition involves matching a query image taken without the subject's cooperation, and typically exhibits greater variations in confounding factors such as pose, illumination, expression, resolution, and occlusion [12].

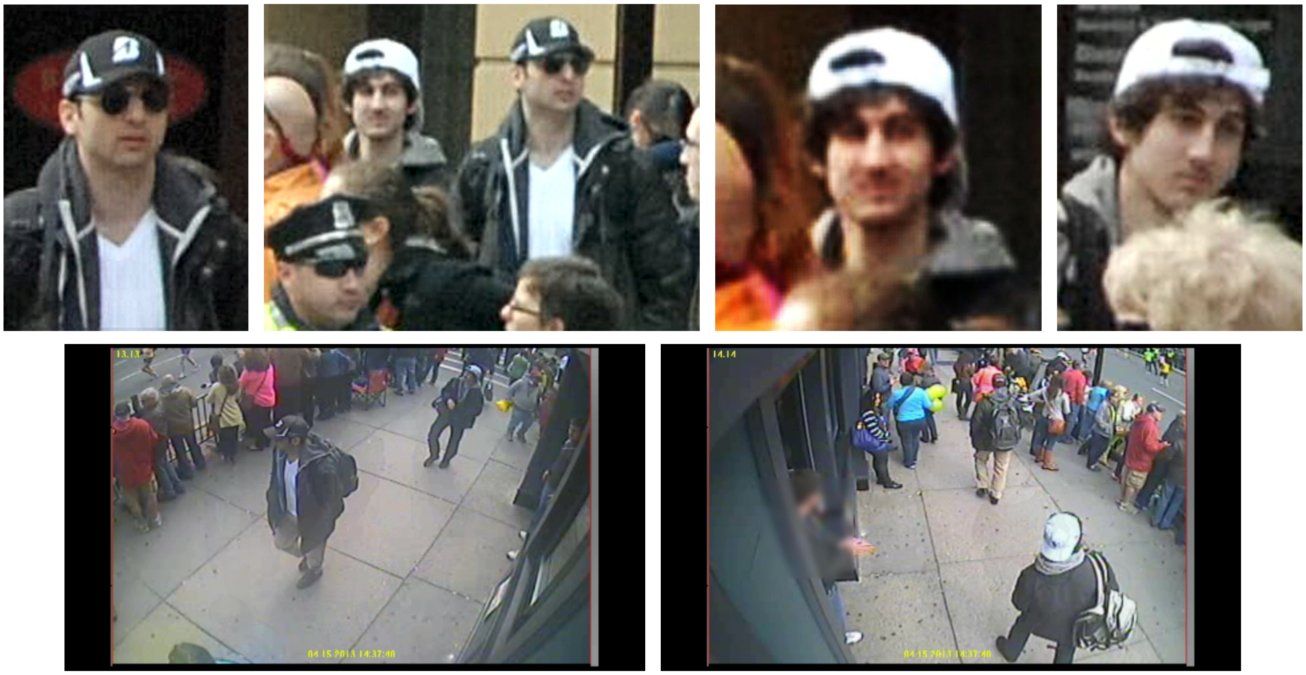

Figure 1: Facial images and videos released by the FBI of the two suspects in the Boston Marathon bombings [3]. Suspect 1, Tamerlan Tsarnaev, is wearing a black hat. Suspect 2, Dzhokhar Tsarnaev, is wearing a white hat. The public was asked to help identify these two individuals.

Figure 2: Timeline of events surrounding the Boston Marathon bombings investigation. There was an 88 hour window for opportunity where facial recognition could have assisted in identification of the suspects.

1.1. Similar Events

There have been a number of cases similar to the Boston Marathon bombings where a mature face recognition technology could have assisted law enforcement in identifying suspects. We summarize three such cases below.

On July 7, 2005 four bombs were detonated on the London public transportation system, killing 52 civilians and injuring more than 700 others [10]. Law enforcement was able to leverage over 6,000 hours of CCTV footage to reconstruct the movements of the bombers as they made a reconnaissance ahead of the actual attacks and entered the subway system [10]. To our knowledge, no attempt was made at the time to run automated facial recognition systems on the CCTV footage.

On June 15, 2011 a riot broke out in downtown Vancouver, injuring 140 people, following the loss of the Vancouver Canucks in the Stanley Cup finals. The Integrated Riot Investigation Team (IRIT) collected approximately 15,000 images and nearly 3,000 videos following the event [11]. In an unprecedented move, the IRIT launched a website showing faces of individuals who participated in the riot, and asked the public to help identify those involved [1]. As of this writing, 13.9 million images have been viewed leading to charges against 221 suspects. An attempt to use automated facial recognition to help identify the rioters was rejected due to privacy violations [7].

Between the 6th and 10th of August 2011, riots and disturbances broke out in London following a peaceful protest in response to the police handling of the shooting of Mark Duggan [20]. Law enforcement published photographs of rioters caught on CCTV cameras or news footage with the hope that witnesses would come forward to identify the suspects. Automated facial recognition technology was largely unsuccessful in providing positive identifications, including one notable attempt by amateurs leveraging *Face.com* [23].

2. Experimental Setup

We simulate the automated facial recognition scenario presented by the Boston Marathon bombings using three state-of-the-art commercial face recognition systems, and images published by law enforcement and news agencies. The following sections describe how the dataset and matchers were selected.

Figure 3 shows the five probe (or query) images considered in our experiments, cropped from photographs in Figure 1. No preprocessing was performed prior to enrollment, though probes *2a* and *2b* appear to originate from the same image, suggesting *2b* may have been modified before it was published. Given the difficulty of automatic face detection, quality estimation, tracking, and activity recognition in uncontrolled environments, we assume that these face images were extracted manually by law enforcement officials.

2.1. Dataset

Figure 4 shows the six gallery images of the two suspects considered in this experiment. Image *1x* is a booking photo of the first suspect from a 2009 arrest in Cambridge, Massachusetts [4]; *1y* is a photo of the first suspect accepting a trophy for winning the 2010 New England Golden Gloves Championship in Lowell, Massachusetts [21]; and *1z* depicts the suspect following a 2009 boxing match in Salt Lake City, Utah [15]. Image *2x* of the second suspect was released by the FBI following his identification but prior to his capture [6]; *2y* is the suspect posing in a high school graduation photo [24]; and *2z* is an unspecified photograph released in a "wanted" flyer by the Boston Regional Intelligence Center [18].

The six gallery images were added to a background set of one million mugshot photographs from the Pinellas County Sheriff's Office (PCSO). The mugshots were acquired in the public domain through Florida's "Sunshine" laws. Figure 5 shows the demographic makeup of the PCSO dataset, and Figure 6 provides some example photographs.

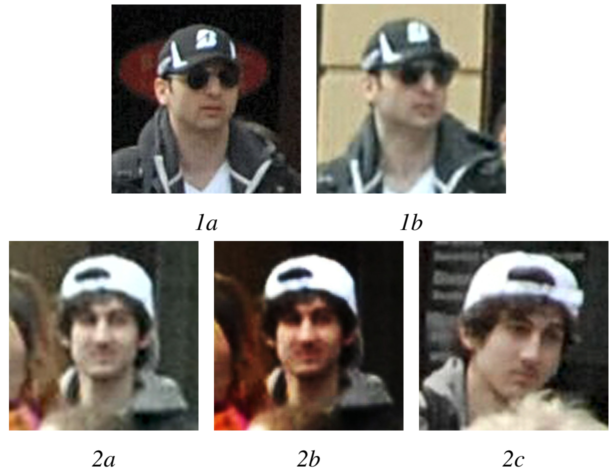

Figure 3: Selected probe images of the two suspects from media released by the FBI [3]. Face images *1a* and *1b* are the two probe images used for Suspect 1. Face images *2a*, *2b* and *2c* are the three probe images used for Suspect 2.

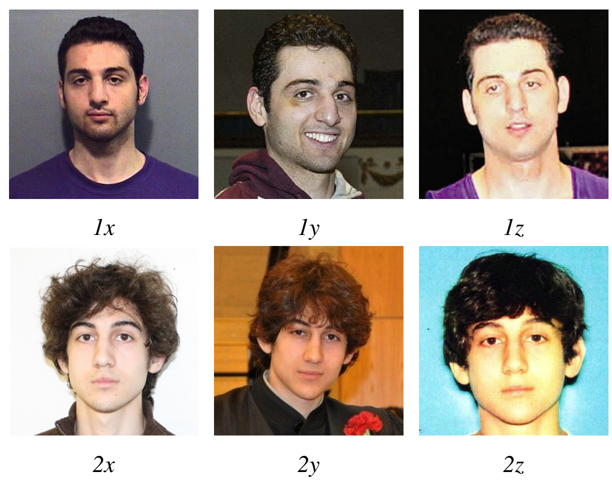

Figure 4: Selected gallery images of the two suspects from varying sources [4, 6, 15, 18, 21, 24] released following the identification of the suspects. Face images *1x*, *1y* and *1z* are the three gallery images of Suspect 1. Face images *2x*, *2y* and *2z* are the three gallery images of Suspect 2.

2.2. Matchers

Two state-of-the-art commercial matchers, NEC NeoFace 3.1[2] and Cognitec FaceVACS 8.6[3], were chosen based on their top performances in the National Institute of Standards and Technology (NIST) Multiple Biometrics Evalua-

[2] www.nec.com/en/global/solutions/security/products/face_recognition.html
[3] www.cognitec-systems.de/FaceVACS-SDK.19.0.html

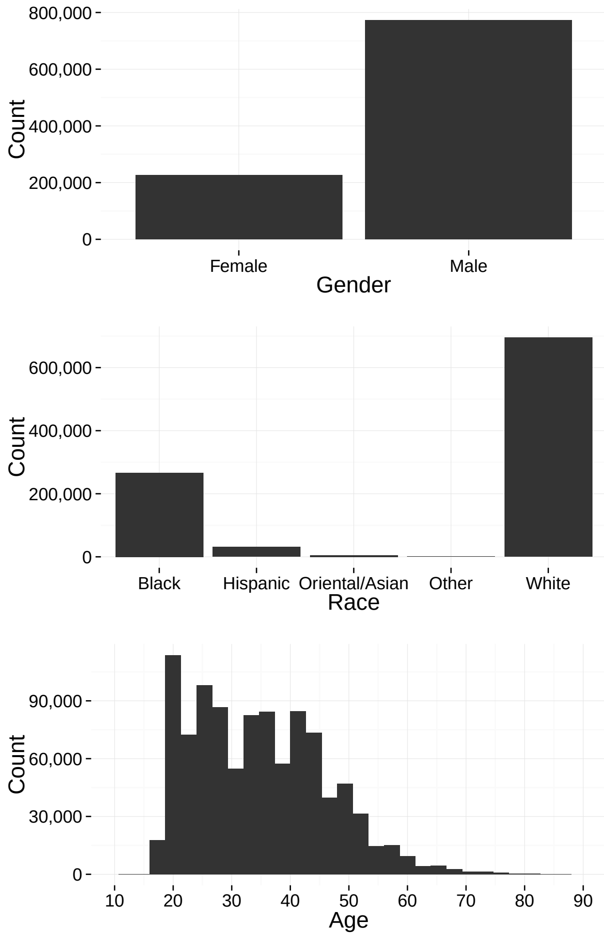

Figure 5: Demographic makeup of the one million PCSO mugshots used as gallery images.

Figure 6: Examples of the one million PCSO mugshots used as gallery images.

tion (MBE) 2010 test. Against a dataset of 1.6 million law enforcement booking images, NeoFace placed first with a rank-one retrieval rate of 92% and FaceVACS placed third with a rank-one retrieval rate of 83% [9]. NeoFace also exhibited notably strong invariance to yaw and elapsed time in this study, and inter-eye distance and compression in [22]. PittPatt 5.2.2[4] was also selected for its prevelant use within the law enforement community and superior performance on non-frontal facial images. In general, matchers were run with their most permissive settings in order to enroll the unconstrained query images, though no other parameter tuning was conducted.

3. Face Matching Results

Three separate experiments measuring ranked retrieval rate were conducted to assess the performance of the face matchers in different scenarios.

3.1. Blind Search

In the blind search, each probe is compared against all gallery images without utilizing the demographic information (e.g., gender, ethnicity and age) associated with gallery faces. Table 1 shows the retrieval rankings for each probe. Probes *1a* and *1b* needed manual assignment of eye locations in order to enroll in FaceVACS, and could not be enrolled in PittPatt as its SDK does not allow for manual eye localization. NeoFace outperforms FaceVACS as well as PittPatt on all probe images in our experiments. PittPatt performs better than FaceVACS on probes *2a*, *2b* and *2c*.

The NIST MBE 2010 offers some insight into the different engineering trade-offs made by NeoFace and FaceVACS, and could explain the disparity in performance observed here. FaceVACS may leverage micro facial features including scars, facial marks and other Level 3 features [14], which would explain its superior performance at very low false accept rates [9] and inferior performance on highly compressed images [22]. Inversely, NeoFace may leverage a more holistic face representation using Level 1 and Level 2 features [14], which would explain its inferior performance at very low false accept rates [9] and superior performance on highly compressed images [22]. As a result, further discussion will focus primarily on NeoFace due to its higher accuracy in our experiments.

Probes for the younger brother, Dzhokhar Tsarnaev exhibited notably better retrieval rates than probes for Tamer-

[4] Acquired by Google

NeoFace 3.1	1x	1y	1z
1a	116,342	**12,446**	87,501
1b	471,165	438,207	**236,343**
	2x	2y	2z
2a	**213**	308	3,353
2b	7,460	**260**	34,013
2c	1,869	**1**	12,622
FaceVACS 8.6	1x	1y	1z
1a	800,596	559,057	527,252
1b	853,906	663,030	759,100
	2x	2y	2z
2a	51,143	306,802	283,932
2b	882,467	864,931	737,555
2c	139,699	206,676	403,867
PittPatt 5.2.2	2x	2y	2z
2a	14,965	5,556	7,470
2b	997,871	9,002	5,779
2c	139	636	39,943

Table 1: Blind (exhaustive) search rankings. Each row contains the ranks at which the true mated gallery images were returned for a given probe. Bold numbers indicate the lowest rank true mate returned for each probe.

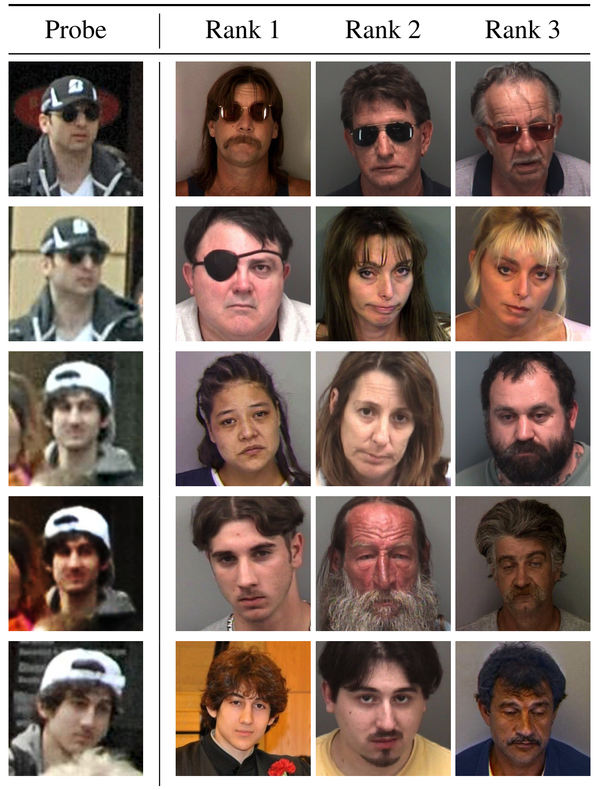

Figure 7: Top three retrievals in a blind search with NeoFace 3.1.

lan Tsarnaev whose face was occluded by sunglasses. Probe *2b*, which appears to be an "enhanced" version of *2a*, generally resulted in lower matching accuracy. For the most part, gallery images *1y* and *2y* were retrieved at the lowest ranks, with pose consistency between gallery and probe seeming to be the crucial factor. Notably, probe *2c* returned gallery image *2y* as a rank-one hit.

Figures 7, 8, and 9 show the top three returns of each probe for NeoFace 3.1, FaceVACS 8.6 and PittPatt 5.2.2, respectively. The sunglasses worn by the older brother, Tamerlan Tsarnaev appear to have significantly degraded his top matches. General inconsistencies between the demographics of each probe and its top returns from the gallery suggest that demographic filtering would improve the accuracy.

3.2. Filtered Search

In the filtered search, each probe is only compared against gallery images with similar demographic data [13]. For Suspect 1 (white, male, 20 to 30 years old) and Suspect 2 (white, male, 15 to 25 years old), filtering reduced the size of the PCSO background gallery from one million to 174,718 and 131,462 images, respectively.

Table 2 shows the gallery retrieval rankings for each probe, and Figures 10, 11, and 12 show the top three returns of each probe for NeoFace 3.1, FaceVACS 8.6 and PittPatt 5.2.2, respectively. Demographic filtering substantially improves retrieval rankings compared to the blind search, with an improvement generally proportional to the reduction in gallery size.

3.3. Fused Search

In the fused search, match scores using different probe images of the same suspect are summed up without weighting before ranking the gallery images. Table 3 shows the gallery retrieval rankings for fused probes with and without demographic filtering. In general, fusion improves retrieval rates for gallery images ranked similarly by each probes, but degrades performance for gallery images ranked differently across the fused probes.

4. Summary

While the Boston Marathon bombings case offers only a small number of published face images for automatic matching, we believe there is still valuable insight to be gained from an interpretation of the results. To begin with, not all commercial face recognition systems appear equally

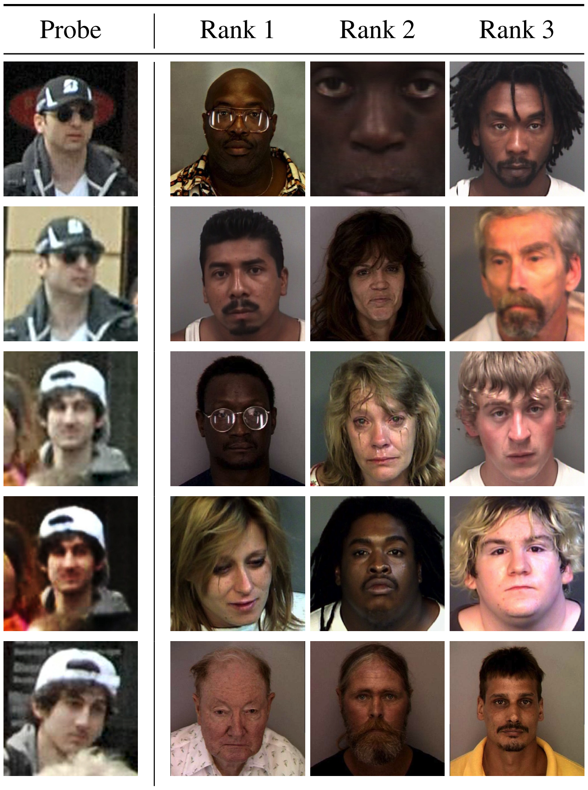

Figure 8: Top three retrievals in a blind search with Face-VACS 8.6.

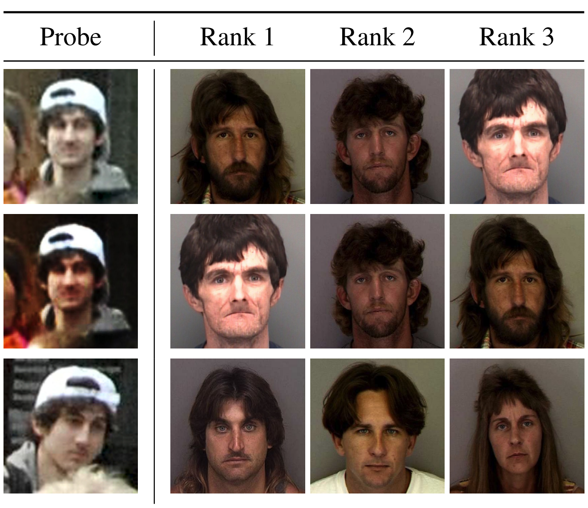

Figure 9: Top three retrievals in a blind search with PittPatt 5.2.2. PittPatt could not enroll probes *1a* and *1b*.

well suited for the unconstrained face recognition scenario. While both NeoFace and FaceVACS demonstrate high accuracy for conventional mugshot-to-mugshot recognition, the NeoFace algorithm exhibits much better performance in our experiments on unconstrained face recognition. PittPatt

NeoFace 3.1	1x	1y	1z
1a	17,858	**1,746**	13,253
1b	83,651	78,024	**42,827**
	2x	2y	2z
2a	**19**	29	253
2b	761	**30**	3541
2c	267	**1**	1703
FaceVACS 8.6	1x	1y	1z
1a	130,944	84,974	79,210
1b	152,988	121,607	137,442
	2x	2y	2z
2a	109,150	13,254	69,327
2b	113,728	111,172	93,115
2c	16,735	25,130	51,069
PittPatt 5.2.2	2x	2y	2z
2a	2,051	753	1,012
2b	131,355	1,339	856
2c	28	139	7,803

Table 2: Filtered search retrieval rankings. Each row contains the ranks at which the true mated gallery images were returned for a given probe. Bold numbers indicate the lowest rank true mate returned for each probe.

performed somewhere in between the NeoFace and Face-VACS. We hope to extend these results with other commercial systems.

Even with NeoFace, the matching accuracy is likely not yet accurate enough for "lights out" deployment in law enforcement applications. More progress must be made in overcoming challenges such as pose, resolution, and occlusion in order to increase the utility of unconstrained facial imagery. Still, with demographic filtering, multiple probes, and a human in the loop, state-of-the-art face matchers can potentially assist law enforcement in apprehending suspects in a timely fashion.

The notable rank-one hit for Dzhokhar Tsarnaev is an illustrative example of this potential. However, the hit was against a graduation photograph posted on Facebook with similar pose, and not a conventional mugshot. This demonstrates the potential value in searching multiple face databases.

References

[1] J. Chu. VPD annual report and riot website launch. Vancouver Police Department, August 30 2011.

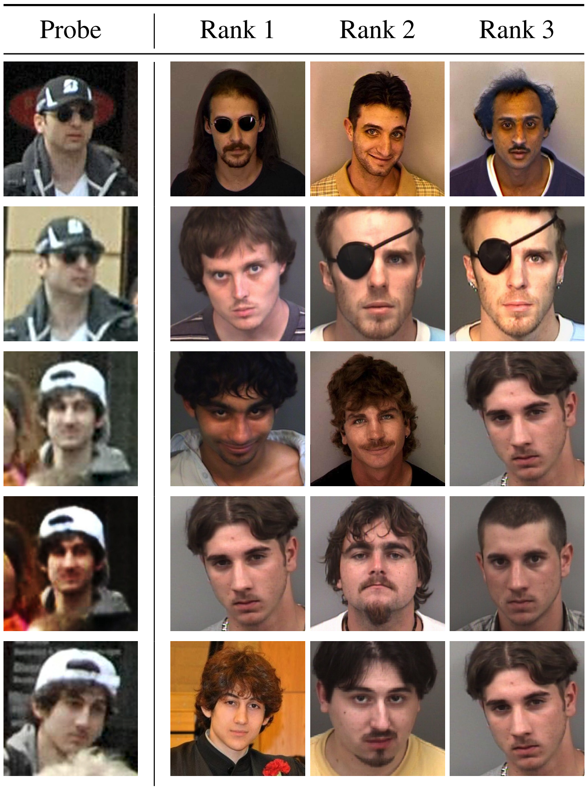

Figure 10: Top three retrievals in a demographically filtered search with NeoFace 3.1.

http://mediareleases.vpd.ca/2011/08/30/vpd-annual-report-and-riot-website-launch/.

[2] G. Comcowich. Remarks of special agent in charge Richard DesLauriers at press conference on bombing investigation. FBI Boston, April 16, 2013. http://www.fbi.gov/boston/press-releases/2013/remarks-of-special-agent-in-charge-richard-deslauriers-at-press-conference-on-bombing-investigation.

[3] G. Comcowich. Remarks of special agent in charge Richard DesLauriers at press conference on bombing investigation. FBI Boston, April 18, 2013. http://www.fbi.gov/boston/press-releases/2013/remarks-of-special-agent-in-charge-richard-deslauriers-at-press-conference-on-bombing-investigation-1.

[4] T. Connor. Funeral director in Boston bombing case used to serving the unwanted. U.S. News, May 6, 2013. http://usnews.nbcnews.com/_news/2013/05/06/18086503-funeral-director-in-boston-bombing-case-used-to-serving-the-unwanted.

[5] T. De Chant. The limits of facial recognition. PBS NOVA, April 26, 2013. http://www.pbs.org/wgbh/nova/next/tech/the-limits-of-facial-recognition/.

[6] FBI. Updated photo of suspect 2 released. FBI, April 19, 2013. http://www.fbi.gov/news/updates-on-investigation-into-multiple-explosions-in-boston.

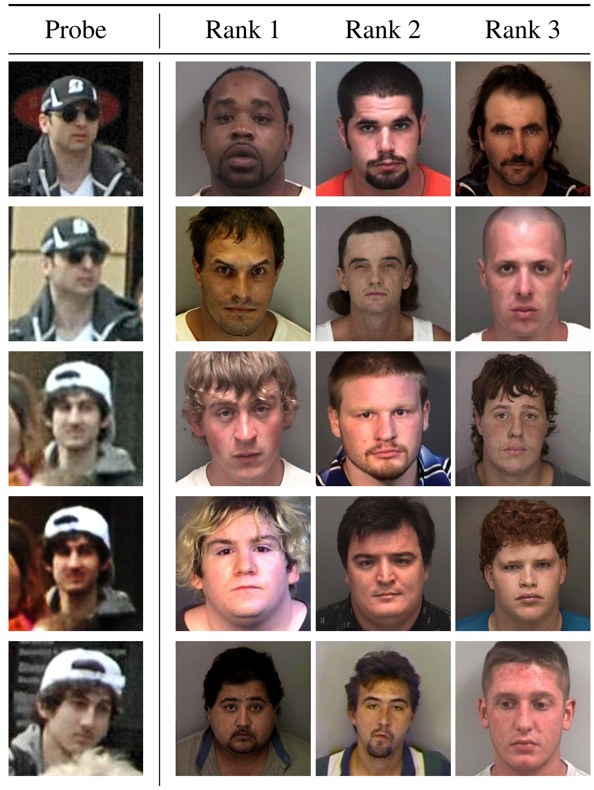

Figure 11: Top three retrievals in a demographically filtered search with FaceVACS 8.6.

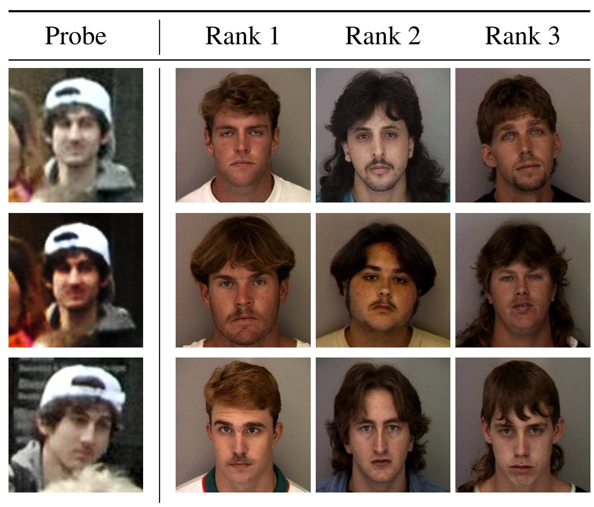

Figure 12: Top three retrievals in a demographically filtered search with PittPatt 5.2.2. PittPatt could not enroll probes *1a* and *1b*.

[7] J. Fowlie. Court order required to use facial recognition to identify Stanley Cup rioters. The Vancouver Sun, February 17, 2012. http://www.vancouversun.com/news/Court+order+required+

NeoFace 3.1	Filtered	1x	1y	1z
1a+1b	No	217,761	**48,982**	122,325
1a+1b	Yes	29,143	**6,591**	16,453
		2x	2y	2z
2a+2c	No	74	**3**	1,798
2a+2c	Yes	15	**2**	179
FaceVACS 8.6	Filtered	1x	1y	1z
1a+1b	No	940,378	741,564	764,526
1a+1b	Yes	163,415	126,666	130,861
		2x	2y	2z
2a+2c	No	327,544	128,813	544,254
2a+2c	Yes	40,576	15,094	70,751
PittPatt 5.2.2	Filtered	2x	2y	2z
2a+2c	No	493	527	10,048
2a+2c	Yes	69	75	1,660

Table 3: Score level sum fusion retrieval ranks with and without demographic filtering. Each row contains the ranks at which the true mated gallery images were returned for a given probe. Bold numbers indicate the lowest rank true mate returned for each probe.

facial+recognition+identify+Stanley+rioters/6163995/story.html.

[8] S. Gallagher. Why facial recognition tech failed in the Boston bombing manhunt. Ars Technica, May 7, 2013. http://arstechnica.com/information-technology/2013/05/why-facial-recognition-tech-failed-in-the-boston-bombing-manhunt/.

[9] P. Grother, G. Quinn, and J. Phillips. Report on the evaluation of 2d still-image face recognition algorithms. NIST Interagency Report 7709, June 2010. http://www.nist.gov/manuscript-publication-search.cfm?pub_id=905968.

[10] House of Commons. Report of the official account of the bombings in london on 7th july 2005, May 11, 2006. http://www.official-documents.gov.uk/document/hc0506/hc10/1087/1087.asp.

[11] M. Howell. Stanley Cup riot charges may take another two months. The Vancouver Courier, July 11, 2011. http://www.vancourier.com/news/Stanley+riot+charges+take+another+months/5085446/story .html.

[12] A. K. Jain, B. Klare, and U. Park. Face matching and retrieval in forensics applications. *IEEE MultiMedia*, 19(1):20, 2012.

[13] B. Klare, M. Burge, J. Klontz, R. Vorder Bruegge, and A. Jain. Face recognition performance: Role of demographic information. *IEEE Trans. on Information Forensics and Security*, 7(6):1789–1801, 2012.

[14] B. Klare and A. K. Jain. On a taxonomy of facial features. In *IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), 2010*, pages 1–8, 2010.

[15] H. Maass. 10 things you need to know today: May 6, 2013. The Week, May 6, 2013. http://theweek.com/article/index/243734/10-things-you-need-to-know-today-may-6-2013.

[16] S. Malone and G. McCool. Boston officials say 264 injured in marathon bombing. Reuters, April 23, 2013. http://www.reuters.com/article/2013/04/23/us-usa-explosions-boston-injuries-idUSBRE93M0LW20130423.

[17] T. McLaughlin, R. Kerber, S. Malone, S. Herbst-Bayliss, F. McGurty, and T. Dobbyn. A shaken Boston mostly gets back to work; 12-block crime scene. Reuters, April 16, 2013. http://www.reuters.com/article/2013/04/16/us-usa-explosions-boston-workers-idUSBRE93F10X20130416.

[18] M. Memmott and E. Peralta. 'the hunt is over:' police apprehend marathon bombing suspect. NPR, April 19, 2013. http://www.npr.org/blogs/thetwo-way/2013/04/19/177885868/shots-explosions-heard-as-boston-manhunt-continues.

[19] D. Montgomery, S. Horwitz, and M. Fisher. Police, citizens and technology factor into Boston bombing probe. The Washington Post, April 20, 2013. http://articles.washingtonpost.com/2013-04-20/world/38693691_1_boston-marathon-finish-line-images.

[20] G. Morrell, S. Scott, D. McNeish, and S. Webster. The August riots in England. National Centre for Social Research, November 2011. http://www.natcen.ac.uk/study/the-august-riots-in-england-.

[21] E. Ortiz. Dead Boston bombing suspect Tamerlan Tsarnaev lost American friend to grisly murder two years ago. New York Daily News, April 20, 2013. http://www.nydailynews.com/news/crime/tamerlan-tsarnaev-lost-american-friend-grisly-murder-article-1.1322990.

[22] G. Quinn and P. Grother. Performance of face recognition algorithms on compressed images. NIST Interagency Report 7830, December 2011. http://www.nist.gov/manuscript-publication-search.cfm?pub_id=908515.

[23] A. Saenz. Scotland Yard using facial recognition to find rioters - but tech isn't up to the task. SingularityHUB, August 20, 2011. http://singularityhub.com/2011/08/20/scotland-yard-using-facial-recognition-to-find-rioters-

[24] Staff. Boston bomb suspect Dzokhar Tsarnaev hunted by SWAT teams after police kill boxer brother Tamerlan in running gun battle. news grio, April 19, 2013. http://www.newsgrio.com/latest-news/34725-boston-bomb-suspect-dzokhar-tsarnaev-hunted-by-swat-teams-after-police-kill-boxer-brother-tamerlan-in-running-gun-battle.html.

[25] M. Stroud. In Boston bombing, flood of digital evidence is a blessing and a curse. The Verge, April 16, 2013. http://www.theverge.com/2013/4/16/4230820/in-boston-bombing-flood-of-digital-evidence-is-a-blessing-and-a-curse.