The Challenge of Face Recognition from Digital Point-and-Shoot Cameras

J. Ross Beveridge* P. Jonathon Phillips† David Bolme‡ Bruce A. Draper*,
Geof H. Givens§ Yui Man Lui* Mohammad Nayeem Teli* Hao Zhang*
W. Todd Scruggs¶ Kevin W. Bowyer‖ Patrick J. Flynn‖ Su Cheng†

Abstract

Inexpensive "point-and-shoot" camera technology has combined with social network technology to give the general population a motivation to use face recognition technology. Users expect a lot; they want to snap pictures, shoot videos, upload, and have their friends, family and acquaintances more-or-less automatically recognized. Despite the apparent simplicity of the problem, face recognition in this context is hard. Roughly speaking, failure rates in the 4 to 8 out of 10 range are common. In contrast, error rates drop to roughly 1 in 1,000 for well controlled imagery. To spur advancement in face and person recognition this paper introduces the Point and Shoot Face Recognition Challenge (PaSC). The challenge includes 9,376 still images of 293 people balanced with respect to distance to the camera, alternative sensors, frontal versus not-frontal views, and varying location. There are also 2,802 videos for 265 people: a subset of the 293. Verification results are presented for public baseline algorithms and a commercial algorithm for three cases: comparing still images to still images, videos to videos, and still images to videos.

*J. R. Beveridge, B. A. Draper, Y-M Lui, M. N. Teli, and H. Zhang are with the Department of Computer Science, Colorado State U., Fort Collins, CO 46556, USA.

†P. J. Phillips and S. Cheng are with the National Institute of Standards and Technology, 100 Bureau Dr., MS 8940 Gaithersburg MD 20899, USA

‡D. S. Bolme is with the Oak Ridge National Lab., Knoxville TN

§G. Givens is with the Department of Statistics, Colorado State U., Fort Collins, CO 46556, USA.

¶W. T. Scruggs is with Science Applications International Corporation, 14668 Lee Road, Room 4088, Chantilly, VA 20151.

‖K. W. Bowyer and P. J. Flynn are with the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA.

[1]"Work at CSU was supported by the Technical Support Working Group (TSWG) and IARPA. PJP and SC were funded in part by the Federal Bureau of Investigation. DB was at CSU when this work was performed. The identification of any commercial product or trade name does not imply endorsement or recommendation by Colorado State U., Oak Ridge National Laboratory, NIST, SAIC, or U of Notre Dame. The collection of the PaSC data set was funded under IARPA's BEST program.

1. Introduction

The difficulty of automatic face recognition grows dramatically as constraints on imaging conditions are relaxed. Under the most controlled conditions, defined as frontal face images taken in mobile studio or mugshot environments, the Multiple Biometric Evaluation (MBE) 2010 reported a verification rate of 0.997 at a false accept rate (FAR) of 1 in a 1,000 [7], as shown in Figure 1. When the illumination conditions are relaxed and frontal face images are acquired with a digital single-lens reflex camera under natural indoor and outdoor lighting conditions, the corresponding verification rate drops to 0.80 [15]. This is the overall performance for the three partitions in the Good, Bad, & Ugly (GBU) data set. The Labeled Faces in the Wild (LFW) challenge problem contains images of celebrities and famous people collected off the web [8]. These images were originally acquired by photo-journalists and curated prior to posting on the web. To date the best performance is a verification rate of 0.54 at a FAR = 0.001 [2].

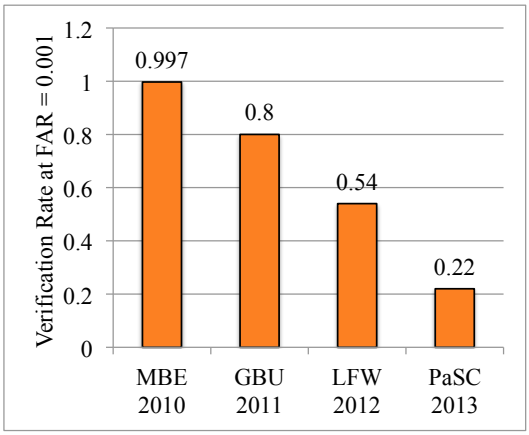


Figure 1. Performance progressively drops when shifting from controlled scenarios to uncontrolled point-and-shoot conditions.

Today the majority of face images acquired world wide are taken by amateurs using point-and-shoot cameras and

cell phones. The resulting images include complications rare in the previous scenarios, including poor staging, blur, over and under exposure, and compression artifacts. This is despite mega-pixel counts of 12 to 14 on some of these cameras.

To call attention to these issues, we introduce the Point-and-Shoot Challenge (PaSC). The last bar in Figure 1 shows the verification rate on PaSC still images using the SDK 5.2.2 version of an algorithm developed by Pittsburgh Pattern Recognition (PittPatt)[1].

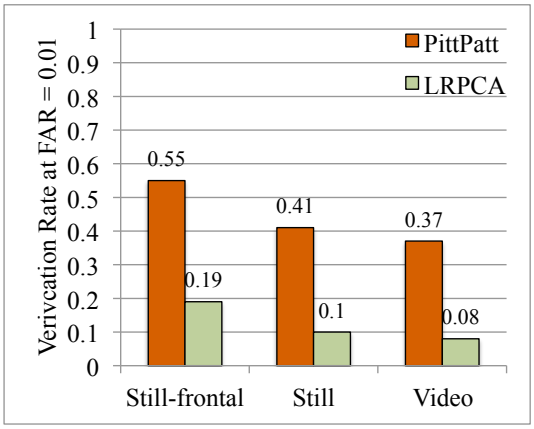


Figure 2. Verification rates for frontal still images in PaSC (left), frontal and non-frontal still images in PaSC (middle), and video-to-video comparisons in PaSC. Performance is reported for the PittPatt SDK and local region principal component analysis (LRPCA) algorithm [15]. LRPCA is an open source baseline. Note that because of the difficulty of the problem, verification rates are provided at a false accept rate (FAR) of 0.01, instead of the tradition FAR of 0.001.

The PaSC includes both still images and videos. Figure 2 summarizes verification performance on frontal still images, a mix of frontal and non-frontal views, and video-to-video comparisons. In acknowledgement of the increased difficulty in the PaSC, verification rates in Figure 2 are reported at FAR = 0.01 instead of 0.001. The inclusion of still and video provides a powerful opportunity to explore the relative value of each.

2. The PaSC Overview

There are $9,376$ images of 293 people in the still portion of the PaSC. Image collection was carried out according to an experiment design that systematically varied sensor, location, pose and distance from the camera. The PaSC also includes $2,802$ videos of 265 people carrying out simple actions. These people are a subset of the 293 people in the still image portion of the PaSC. This design facilitates statistical analysis of how the factors just enumerated influence still and video face recognition.

Figure 3. Examples of images in the PaSC taken during four sessions. Note that locations were varied between sessions, while sensor, distance to camera and pose were varied within sessions.

Figure 4. Cropped face images extracted from still images in the PaSC. These images demonstrate some of the complications that arise in point-and-shoot images, lighting, motion blur and poor focus.

To provide more background on the data collection process, Figures 3 shows images from four different sessions, while Figure 4 shows a sample of detected and cropped faces. Figure 3 illustrates variation in location, sensor, distance to camera, and pose (frontal vs. non-frontal). While Figure 3 shows thumbnails, the actual images are approximately 3000 by 4000 pixels: exact size varies by sensor. Median face size expressed as pixels between the eyes is 68; quartile boundaries are 15, 54, 68, 116 and 476 (including min and max). The closeups in Figure 4 illustrate some of the aspects of this data that make it challenging.

Figure 5 shows 4 frames from one video taken with the highest resolution handheld video camera (1280 x 720). Subjects are instructed to carry out actions in order to capture a wider variety of views. Another motivation is to discourage scenarios where a person looks directly into the video camera for a prolonged time, reducing the task to

[1]The PittPatt SDK was used in the experiments because it was available under a U.S. Government use license.

Figure 5. Four snapshots from one video showing a subject carrying out an action, in this case blowing bubbles. Frame captures are down sampled by half in this figure to better fit as a four panel figure.

frontal image recognition with multiple stills. Different actions were scripted for different locations and days. In all cases, however, the videos show a person entering the scene relatively far from the camera, carrying out an action, and then leaving the field of view. Typically the person is relatively close to the camera when they leave the field of view, but they do not look directly at the camera.

The data for this challenge problem is available upon request through the following website (url omitted in review manuscript). A separate website, (url omitted), supports the challenge in the following ways. First, it has a download facility where anyone may download software and metadata associated with the challenge. This software includes the baseline algorithms described below as well as code for generating similarity matrices and associated ROC curves. The site is curated and has facilities for researchers to register as active contributors and upload results in a manner that coordinates the tracking of progress on the PaSC.

3. Related Work

The FERET evaluation [14] was the first significant effort in face recognition to distribute a common data set along with an established standard protocol. Since then a variety of data sets, competitions, evaluations, and challenge problems have contributed to the face recognition field. Here we highlight a few.

The CMU PIE face database [5] was collected in such a way as to support excellent empirical explorations of controlled interactions between Illumination and Pose. The Multi-PIE face database [6] released in 2010 includes 337 people and, like its predecessor, supports empirical analysis with densely sampled and controlled variations in illumination and pose. The Face Recognition Grand Challenge [16] Experiment 4 matched indoor controlled images to indoor and outdoor uncontrolled lighting images and constituted a major new challenge.

The XM2VTS and BANCA Databases [1] were each released with associated evaluation protocols and competitions were organized around each [10, 11]. The European BioSecure project represents a major coordinated effort to advance the multi-modal biometrics, including face [13]. The associated BioSecure DS2 (Access Control) evaluation campaign [17] emphasized fusion along with the use of biometric quality measures.

A desire to move away from controlled image acquisition scenarios is well expressed in the Labeled Faces in the Wild [8] dataset. Two notable aspects of LFW are the shift to images of opportunity, for LFW images on the web, along with a well coordinated and updated website that curates current performance results. Face detection also grows more difficult in less controlled scenarios, and the recent Face Detection on Hard Datasets Competition [12] brought together many groups in a joint effort.

For more background on video face recognition approaches there is an excellent recent survey [3]. Open datasets are emerging, for example the YouTube Faces dataset consisting of 3425 videos of 1595 different people[19] collected in the spirit of LFW and adopting a similar ten-fold, cross validation framework common in machine learning but not in biometrics. Another previous video challenge problem is the Video Challenge portion of the Multiple Biometric Grand Challenge (MBGC) [2] which, like this challenge, involved people carrying out activities.

4. Protocol

To establish a basis for comparison, and in keeping with the protocol used in previous challenge problems [7, 15], still face recognition algorithms must compute a similarity score for all pairs of images obtained by matching images in a target set to images in a query set. The video portion of the challenge is organized the same way, algorithms compare two videos. A comparison between two videos results in a single similarity score. In all cases, the resulting similarity matrix becomes the basis for subsequent analysis, enabling performance to be expressed in terms of an ROC curve.

There is a further restriction in the protocol which is important to avoid overly optimistic and misleading results. The similarity score $S(q,t)$ returned for a query-target pair (q,t) must not change in response to changes in either the query set Q or the target set T that q and t were drawn from. There are at least two common ways of violating this prohibition that boost performance in experiments:

- Training on images, or people, in T (or Q) is a violation of the PaSC protocol.
- Adjusting similarity scores based on comparisons with other images in either T or Q is a violation of the PaSC protocol.

[2] http://www.nist.gov/itl/iad/ig/focs.cfm

Either or both of these practices will improve verification rates. However, doing so comes at the expense of generality and amounts to addressing a more limited question: "How well does an algorithm perform when tuned in advance for a closed set of known people?"

5. Data Collection Goal and Design

Two goals motivate the PaSC. The first is to create a dataset and challenge problem that will encourage the face recognition community to develop better algorithms and to overcome many of the challenges associated with point-and-shoot data. To this end the dataset was kept small enough for research systems while focusing algorithm development key types of variation:

Location: Still mages were taken at nine locations, both inside buildings and outdoors. Videos were taken in six locations (inside buildings and outdoors).

Pose: Both looking at the camera and off to the side.

Distance: People both near and far from the camera.

Sensor: Five point-and-shoot still cameras, five hand held video cameras, and one control video camera.

Video: Many video frames in contrast to a single still.

The second goal is to support strong statistical analysis. Consider questions such as: "Does distance from the camera to the person matter compared to the choice of camera?" Statistically meaningful answers depend on a data collection plan that ensures a proper sampling across factors as well as balance between factors. To support this goal, before images were collected, the principal institutions contributing to the PaSC jointly developed a data collection plan. Formally, the collection plan resembles a multiway, split-plot design extended over time[4]. The images were collected to achieve a good sampling of the factors enumerated above.

5.1. Still Image Target and Query Sets

The evaluation protocol for the still image portion of the PaSC compares all images in a query set to all images in a target set. Therefore, assignment of images to the target and query sets is tightly coupled with the data collection plan. To illustrate, in the PaSC all people/subjects contribute an equal number of images so that no one subject has a larger influence on results than any other subject. Specifically, each subject contributes 4 images from each of 8 randomly selected locations to the target and query sets. Further, the target and query sets are structured in a way that precludes same-day comparisons: same-day comparisons are well known to be easy and their presence in a data set undermines its value.

Location influences performance because different locations present different imaging conditions. Image selection also balances with respect to location, with approximately the same number of images acquired at each location. However, comparisons between images collected at the same location but on different days are rare, so same-location pairs are favored in the creation of the target and query sets.

Target and query set construction may be explained in terms of blocks. Each block contains typically 4 images collected of one person at one location [3]. Figure 3 illustrates this block structure for a single person, four different locations, near versus far distance, and two sensors. Within each block, four images are randomly selected from those available such that distance and pose are balanced: one image from each of four conditions close/frontal, close/non-frontal, distant/frontal, and distant/non-frontal. If there were insufficient images for a person and location to balance in this fashion then all images for that subject and location are dropped. Sensor selection was randomized to favor equal representation of sensors across the other three factors.

Table 1 summarizes the PaSC still image data. The size is small enough that most algorithm developers can easily experiment with complex algorithms, but still provides a sufficient number of match scores to support strong statistical analysis.

Table 1. Summary of Still Image PaSC Data.

Number of Subjects	293
Total Images	9,376
Images per Subject	32
Match Scores per Subject	256
Total Scores	21,977,344 (4688 X 4688)
Total Match Scores	75,008 (256 per subject)
Number of Locations	9
Same Location Match Scores	8096 (10.8%)

5.2. Video Target and Query Sets

Fewer total videos were collected and consequently some changes are made to take best advantage of the available videos. For example, instead of having disjoint target and query sets, for video there is one set that serves as both the query and target set. Of course, this use of the same set includes the qualification that no video is ever compared directly to itself.

Another distinction is that every action was filmed by two cameras: a high quality, 1920x1080 pixel, Panasonic camera on a tripod and one of five alternative handheld video cameras. The tripod-based Panasonic data serves as a control. The handheld cameras have resolutions ranging from 640x480 up to 1280x720. The control imagery is available for comparison with still image camera images,

[3] In some cases a subject did not visit enough locations to balance in this manner in which case 8 or 12 images from a given location were chosen.

but should generally not be compared with the handheld video, since it is then possible to compare pairs of videos taken at the same time.

Table 2 summarizes the PaSC video data. The information is largely split between the handheld and the control video, reflecting the cautionary point made above that for each handheld video there is a companion control video taken of the same person at the same time doing the same thing. Experiments should generally be run on either the handheld or the control video: not a combination.

Table 2. Summary of Video PaSC Data.

Number of Subjects	265
Total Videos	2,802
Total Control Videos	1,401
Total Handheld Videos	1,401
Control Videos per Subject	4 to 7
Handheld Videos per Subject	4 to 7
Number of Locations	6

5.3. Still Target versus Video Query

Since the PaSC includes still images and video of the same people, it is an excellent data set to begin exploring recognition performance between modalities: still to video. For example, a person known in advance by a single still image is subsequently seen, or is claimed is be seen, in a video. For such an experiment, the query set consists of the 1,401 handheld (or alternatively control) videos and the target set consists of the still image target set. Under such a test, a recognition algorithm must be able to return a single similarity score when provided a single still image and a video.

6. Supporting Data and Software

The previous sections of this paper describe the goals, data, and protocols of the Point-and-Shoot Challenge. This section describes additional data and software that are optional, but are meant to aid researchers. The most significant support comes in the form of baselines for comparisons. Open source baseline algorithms, Cohort linear discriminant analysis (LDA) and LRPCA, are provided for all three versions of the challenge: still to still, video to video, and still to video. PittPatt face detection results are provide provided for both the still and video imagery. For video, results are provided for every frame. As discussed below, there are errors and faces were not detected in every frame. Making the face detection results available allows researchers without access to face detection algorithms to develop face recognition algorithms on video data.

6.1. Baselines for Still Image Matching

Verification results are presented for three algorithms. Two are open-source baseline algorithms, Cohort LDA [9] and LRPCA [15], and are available through the web. The third algorithm was developed by Pittsburgh Pattern Recognition (PittPatt); the results shown were obtained using SDK 5.2.2.

6.1.1 Face Detection and Localization

Results presented on the PaSC problem should include face detection and localization as part of the face recognition algorithm. In other words, results where faces, and more specifically eyes, are manually found are not of great practical interest. However, eye coordinates generated by the PittPatt algorithm are available with the data set. Full studies presenting results on the PaSC may include supplemental results where automated detection is side-stepped by using the provided eye coordinates when this use of the provided coordinates is clearly indicated.

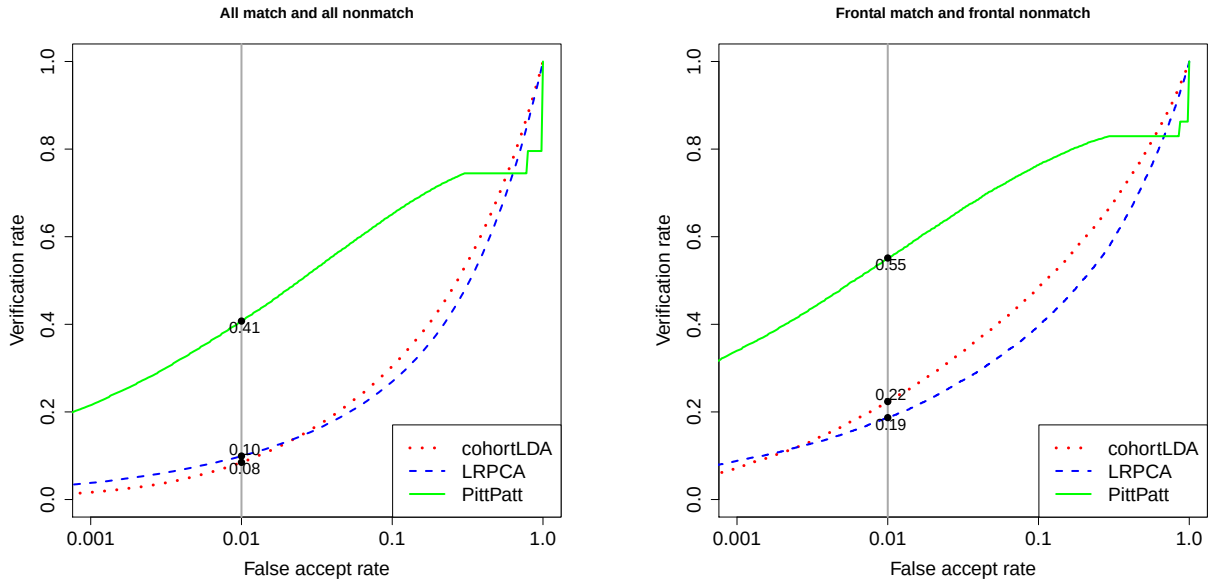
To calibrate the difficulty of face detection and localization on the PaSC data, we have compared the face detector supplied as part of the PittPatt algorithm with the standard OpenCV face detector based upon the work of Viola and Jones [18]. The PittPatt detector failed to locate faces in 7.0% of the PaSC images. In comparison, the OpenCV detector failed to detect the face on 27.6% of the images. We extended the OpenCV algorithm to include information about shoulders as well as faces. The resulting algorithm reduced the error rate to 9.0%, which is not as good as the commercial algorithm but much closer. The code for the extended face detection algorithm is included with software distributed as part of the PaSC.

6.1.2 Verification Performance

Figure 6 shows ROC curves for the PaSC. The Cohort LDA [9] and LRPCA [15] algorithms are baselines distributed as part of the PaSC.

The baseline algorithms are trained using designated training images adhering to the protocol above: both images and people are disjoint relative to the target and query sets. The training image set is provided as part of the PaSC. The PittPatt algorithm came to us pre-trained.

We've tried to avoid subdividing the PaSC into subproblems; doing so weakens the value of a clear singular point of reference. That said, frontal face recognition is a very mature specialization of the problem and we felt it valuable to show in Figure 6b results for the frontal images only. Also note in Figure 6 that verification rates are noted at FAR=0.01. This represents a relaxing of standards relative to the stricter FAR=0.001 standard used in previous



(a) (b)

Figure 6. Still image verification performance of two baseline and one commercial algorithm on the PaSC: a) entire challenge problem, b) frontal face image only portion of the PaSC.

studies [15]. In our opinion this shift is appropriate given the difficulty of the PaSC.

Artifacts are evident in the PittPatt curves starting to the right of FAR=0.1. These are caused by the algorithm setting similarity scores to special constant values, presumably flagging some internal decision made by the algorithm. While the artifacts are prominent and warrant explanation, they are not of practical concern. In practice, verification rates at FARs greater than 0.1 hold little meaning since it is difficult to imagine large scale fielded systems being of value operating at such high false accept rates.

6.2. Video Baseline Results

Several additional levels of complexity arise in addressing the video portion of the challenge. For example, face detection and localization in the handheld video is more difficult than for the still image portion of the PaSC. Another complication concerns how to reduce a comparison between two videos, each running around 5 to 10 seconds in length, to a single similarity score.

6.2.1 Face Detection and Localization in Video

Face detection and localization in the videos is a difficult task. In total, there are 334,879 and 328,967 frames of video in the 1401 control and 1401 handheld videos respectively. The difference is in part due to modest variation in video length. When run on these frames, the OpenCV face detector based upon Viola and Jones [18] frequently fails, finding faces in far fewer than half the total frames of video. The face detection and localization capabilities of the PittPatt software does much better.

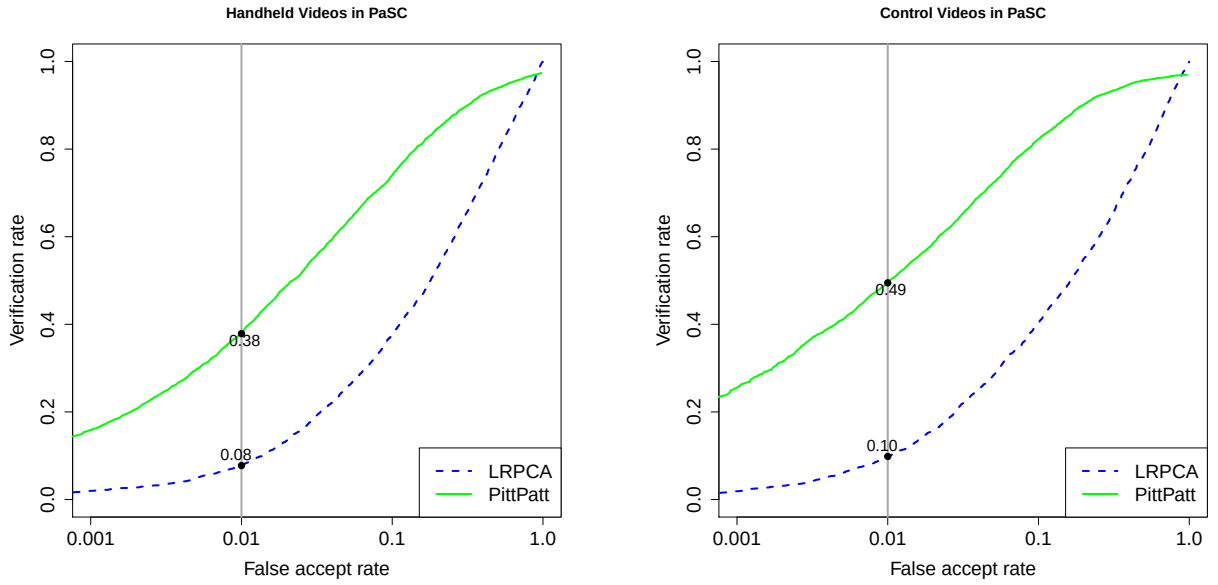
In only 60 out of the 1401 control and 34 out of the 1401 handheld videos did the PittPatt algorithm fail to find a face in any frames. Otherwise, face locations and approximate pose estimates are reported for at lease one frame and typically many frames. To summarize the performance, faces are detected in 121,016 and 127,621 frames of the control and handheld videos respectively. This represents detection in roughly one third of all frames.

To aid research groups that concentrate on the recognition aspects of the challenge, , machine generated face detections will be provided in a csv file containing face location, scale and approximate pose. Algorithms reporting results on the video portion of the PaSC using their own face detection and localization will be reported differently than those that use provided face detection results.

6.2.2 Frame-to-Frame Video Comparisons

Figure 7 shows results comparing videos to videos using an extension of LRPCA and PittPatt algorithms. Specifically, the extension involves comparing all frames with detected faces in one video to all frames with detected faces in another video. The resulting large set of similarity scores is then sorted and a single score is selected based upon rank in the sorted list.

An early discovery in performing this many-to-many



(a) (b)

Figure 7. Video verification performance on the PaSC: a) handheld video, b) on high resolution (1920x1080) control video.

frame comparison followed by a rank based selection of a score is that the maximum score is not always the best. In particular, we examined ROC curves for different choices, including the maximum score, the score 10% into the list of sorted scores, the median score, etc. In the case of LRPCA choosing the max score yielded the best ROC: that is what is shown in Figure 7. For the PittPatt algorithm used in this fashion, the max score yielded a worse ROC than choosing the 90th percentile: that ROC is shown.

One reason CohortLDA results are omitted from Figure 7 is that the ROCs associated with different rank choices in similarity scores was unpredictable and we felt more work is needed to better understand why? In general, what this dependence upon similarity score rank tells us is that more work remains to be done in characterizing how to make such selections and how those selection's consequences propagate into the match and equally importantly the non-match distributions.

6.3. Still Target Versus Video Query Results

Figure 8 presents results where the query set consists of handheld videos and the target set is from the frontal still image to frontal still image test described above. In a fashion similar to the video baseline algorithm, the single still target image is compared to all video frames where a face was detected. The resulting scores are sorted and different algorithm variants arise based upon taking the 'max' score, the '10%' ranked score, etc. It is interesting to note that the verification rate for this experiment is somewhat higher than for the video-to-video experiment.

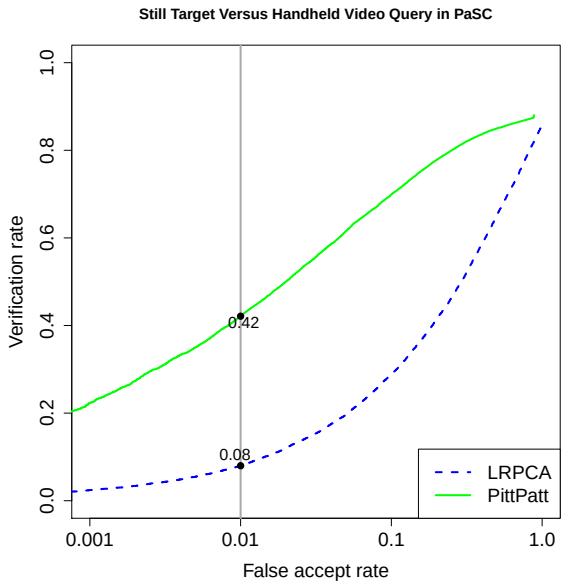


Figure 8. Verification performance comparing still images to videos.

7. Conclusion

The face images, videos, data, and associated metadata for the PaSC are available upon request. The support soft-

ware including the Cohort LDA and LRPCA baseline algorithms and scoring code will be downloadable through the web. We will generate and maintain a curated website where groups working on the PaSC may submit results. Participants may submit performance and quality results. The LFW website clearly demonstrates the value of a common focal point where up-to-date information is available.

Well constructed challenge problems support and promote research over many years, and it is our hope that the PaSC will be the catalyst for substantive and necessary improvements in recognition from point-and-shoot images. Point-and-shoot images introduce new facets to face recognition that we have captured in this challenge. From a covariate perspective, the data collection plan for the PaSC is such that studies carried out on the PaSC can address directly the relative importance of factors such as sensor, location, pose, distance to the camera, etc. The inclusion of video opens up another critical area for research, including fundamental questions with respect to just how much additional benefit, if any, is associated with possessing a video relative to a good single still image.

References

[1] E. Bailly-Bailliére and et al. The BANCA database and evaluation protocol. In *4th International Conference on Audio- and Video-based Biometric Person Authentication*, pages 625–638, 2003. 3

[2] T. Berg and P. Belhumeur. Tom-vs-pete classifiers and identity-preserving alignment for face verification. In *Proceedings of the British Machine Vision Conference*, pages 129.1–129.11. BMVA Press, 2012. 1

[3] R. Chellapa, M. Du, P. Turaga, and S. K. Zhou. *Face Tracking and Recognition in Video*, pages 323 – 351. Springer London, 2011. 3

[4] A. M. Dean and D. Voss. *Design and Analysis of Experiments*. Springer Texts in Statistics. Springer, 1999. 4

[5] R. Gross, S. Baker, I. Matthews, and T. Kanade. Face recognition across pose and illumination. In S. Z. Li and A. K. Jain, editors, *Handbook of Face Recognition*, pages 193–216. Springer-Verlag, June 2004. 3

[6] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing*, 28(5):807 – 813, 2010. 3

[7] P. J. Grother, G. W. Quinn, and P. J. Phillips. MBE 2010: Report on the evaluation of 2D still-image face recognition algorithms. Technical Report NISTIR 7709, National Institute of Standards and Technology, 2010. 1, 3

[8] G. Huang, M. Ramesh, T. Berg, and E. Learned Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report Tech. Rep. 07-49, University of Massachusetts at Amherst, October 2007. 1, 3

[9] Y. M. Lui, D. S. Bolme, P. J. Phillips, J. R. Beveridge, and B. A. Draper. Preliminary studies on the good, the bad, and the ugly face recognition challenge problem. In *CVPR Workshops*, pages 9–16, 2012. 5

[10] J. Matas and et al. Comparison of face verification results on the XM2VTS database. In *Proceedings of the International Conference on Pattern Recognition*, volume 4, pages 4858 – 4853, Barcelona, Spain, September 2000. 3

[11] K. Messer and et al. Face authentication test on the BANCA database. In *Proceedings of the International Conference on Pattern Recognition*, volume 4, pages 523–532, 2004. 3

[12] J. Parris and et al. Face and eye detection on hard datasets. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1 –10, oct. 2011. 3

[13] D. Petrovska-Delacretaz, G. Chollet, and B. Dorizzi. *Guide to Biometric Reference Systems and Performance Evaluation*. Springer, Dordrecht, 2009. 3

[14] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, October 2000. 3

[15] P. J. Phillips and et al. An Introduction to the Good, the Bad, and the Ugly Face Recognition Challenge Problem. In *IEEE Conference on Automatic Face and Gesture Recognition*, pages 346 – 535, March 2011. 1, 2, 3, 5, 6

[16] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005. 3

[17] N. Poh and et al. Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms. *Information Forensics and Security, IEEE Transactions on*, 4(4):849 –866, Dec. 2009. 3

[18] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, May 2004. 5, 6

[19] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 529–534, 2011. 3