The IJCB 2014 PaSC Video Face and Person Recognition Competition

J. Ross Beveridge Hao Zhang
Colorado State University
Fort Collins, CO, USA
ross@cs.colostate.edu

Patrick J. Flynn
University of Notre Dame
Notre Dame, IN, USA

Yooyoung Lee
NIST
Gaithersburg, MD, USA

Venice Erin Liong Jiwen Lu
Advanced Digital Science Center
Singapore

Marcus de Assis Angeloni Tiago de Freitas Pereira
CPqD Telecom & IT Solutions
Campinas, SP, Brasil

Haoxiang Li Gang Hua
Stevens Institute of Technology
Hoboken, NJ, USA

Vitomir Štruc Janez Križaj
University of Ljubljana
Ljubljana, Slovenia

P. Jonathon Phillips
NIST
Gaithersburg, MD, USA

Abstract

The Point-and-Shoot Face Recognition Challenge (PaSC) is a performance evaluation challenge including 1401 videos of 265 people acquired with handheld cameras and depicting people engaged in activities with non-frontal head pose. This report summarizes the results from a competition using this challenge problem. In the Video-to-video Experiment a person in a query video is recognized by comparing the query video to a set of target videos. Both target and query videos are drawn from the same pool of 1401 videos. In the Still-to-video Experiment the person in a query video is to be recognized by comparing the query video to a larger target set consisting of still images. Algorithm performance is characterized by verification rate at a false accept rate of 0.01 *and associated receiver operating characteristic (ROC) curves. Participants were provided eye coordinates for video frames. Results were submitted by 4 institutions: (i) Advanced Digital Science Center, Singapore; (ii) CPqD, Brasil; (iii) Stevens Institute of Technology, USA; and (iv) University of Ljubljana, Slovenia. Most competitors demonstrated video face recognition performance superior to the baseline provided with PaSC. The results represent the best performance to date on the handheld video portion of the PaSC.*

[1]CSU was funded in part by the Department of Defense through the Technical Support Working Group (TSWG). PJP at NIST was supported in part by the FBI.

Table 1. Competition Schedule

November 8, 2013	Competition announced.
February 5, 2014	Participant sign-up cutoff.
March 19, 2014	First round results submitted.
April 22, 2014	Final results submitted.

1. Introduction

This paper summarizes the results of the IJCB 2014 PaSC Video Face and Person Recognition Competition. The competition focused on the problem of recognizing individuals in videos captured using handheld cameras. The individuals in the videos are carrying out an action such as picking up an object or blowing bubbles; they are observed by the camera. Because the subjects are performing an action, the subject's attention is usually directed away from the camera. Complicating factors common in video taken by consumers using devices such as cell phones in everyday settings are emphasized in this competition. Further, videos are taken indoors and outdoors under ambient light and distance from the camera to people shifts over the course of videos. The resulting problem is an excellent complement to the task of recognizing people whose attentions are focused on the camera, and from a face recognition standpoint, this task is very demanding.

The schedule for the competition is shown in Table 1. A summary of the participants submitting final results appears in Table 2. The shortened participant labels are defined to simplify results summaries below. Since the competition is based upon the Point-and-Shoot Face Recognition Challenge [2] (PaSC) [2] data, all participants obtained the associ-

[2]http://www.cs.colostate.edu/~vision/pasc

Table 2. Competition Participants.

Group	Country	Label
Advanced Digital Science Center	Singapore	ADSC
CPqD	Brasil	CPqD
Stevens Institute of Technology	USA	SIT
University of Ljubljana	Slovenia	Ljub

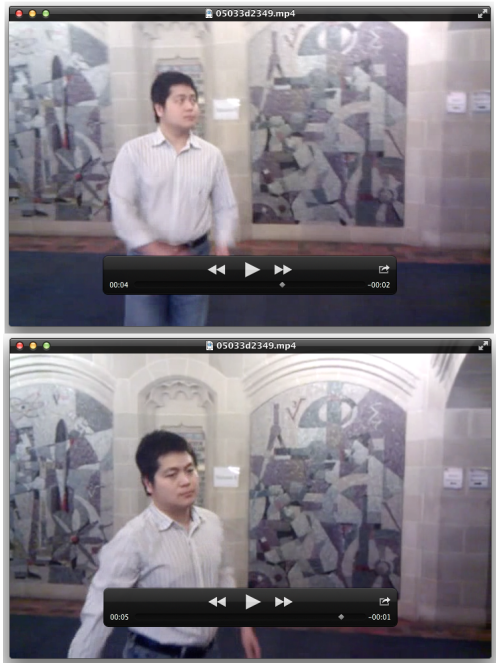

Figure 1. Two full frames of video from the competition highlighting complexity of background, marginal image clarity, and subjects attention being directed away from the camera.

Figure 2. Sampled portions of video frames from PaSC videos indicating some of the situations that make recognition challenging.

ated data from the University of Notre Dame [3].

2. Data Summary and Experiments

This competition used imagery from the handheld video and still image portions of PaSC. The handheld video portion consists of 1401 videos of 265 people acquired at the University of Notre Dame using five different handheld video cameras. The frame sizes for the videos range from 640x480 up to 1280x720. Videos are acquired at six locations, representing a mix of indoor and outdoor settings. Two frames from one video are shown in Figure 1 and a sampling of zoomed portions of frames appears in Figure 2. Figure 3 presents histograms which summarize basic statistics relating to video length, face size and face pose.

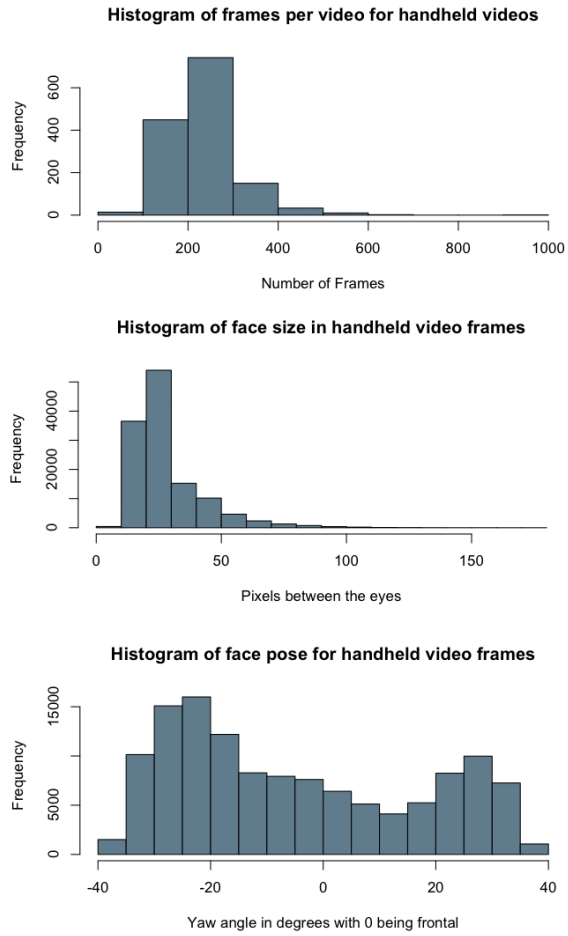

Figure 3. Summary statistics for the handheld video.

The eye distance and pose information presented in Figure 3 was generated automatically using the SDK 5.2.2 version of an algorithm developed by Pittsburgh Pattern Recognition (PittPatt)[4],[5]. The PittPatt algorithm was applied to all frames of video and found faces in $126,220$ out of $328,641$ total frames: the eye distance and pose information is for those frames where faces were detected. The eye coordinates for these $126,220$ frames were supplied to and used by all competition participants.

As Figures 1 and 2 are intended to suggest, and the his-

[3] http://www3.nd.edu/~cvrl/CVRL/Data_Sets.html

[4] The PittPatt SDK was used because it was available under a U.S. Government use license.

[5] The identification of any commercial product or trade name does not imply endorsement or recommendation by NIST.

togram of face pose in Figure 3 shows, in the majority of video frames the face is seen from the side. However, there are points in most videos where close to frontal views are present: note the portion of the pose histogram with yaw angle near zero. Since all results reported here are using the initial face detection results from PittPatt, to some degree the task of selecting favorable video frames has already been done and algorithms are starting from a common point. However, again as the pose and face size histograms in Figure 3 indicate, there is still a great deal of pose and size variation within the video frames.

The still images portion of PaSC consists of images of a superset of the people in the handheld videos. The exact enumeration of videos and still images are provided as part of the Colorado State University (CSU) PaSC Support Software Package[6] in two files:

- pasc_video_handheld.xml: 1401 videos of 265 people.
- pasc_still_target.xml: 4688 images of 293 people.

These files define what are commonly called sigsets, collections of biometric signatures.

Two experiments are defined in the competition:

Video-to-video Experiment The target and query sigsets contained the handheld videos. The task was to verify a claimed identity in the query video by comparing with the associated target video. Since the same 1401 videos served as both the target and query sets, 'same video' comparisons were excluded.

Still-to-video Experiment The target signature set contains still images and the query signature set contains videos (the same set of videos used in the Video-to-video Experiment). The target set includes 4688 images of 293 people.

3. Competition Protocol and Support

The competition was conducted according to the PaSC protocol [2], which requires that the similarity score $s(q, t)$ returned by an algorithm for a query image/video q and target image/video t be unaffected by any other images in the target and query sets. This protocol therefore prohibits training on any images/videos of the people included in the experiments. Similarly, techniques such as cohort normalization using images/videos of the people in the experiments are also prohibited. Finally, to test generalization to new locations, the protocol prohibited training on any imagery collected at the University of Notre Dame during the Spring 2011 semester.

[6] http://www.cs.colostate.edu/~vision/pasc

Training data is supplied as part of the PaSC. This training data was collected by the University of Notre Dame under circumstances broadly similar to the data in the competition. However, the training data comes from collection efforts carried out in semesters different from the competition data, and there are differences. Further, the quantity of video training data was small, only 285 videos. Participants were encouraged to train on other imagery so long as doing so did not violate the restrictions laid down above.

To aid participants with the details associated with running an experiment, the CSU PaSC Support Software Package was made available to all participants. The software includes a baseline local region principle component analysis (LRPCA) video-to-still/video-to-video matching algorithm along with all the surrounding support code needed to encapsulate an experiment and carry it through to the stage of writing out a similarity matrix. Participants were required to submit full similarity matrices along with summary receiver operating characteristic (ROC) curves. Curves generated from the similarity matrices at CSU were returned to participants and checked against their own to test for consistency.

4. Summary of Participants' Approaches

Below are short descriptions of the approaches taken by each of the four participants. As a point of comparison, the results of these four approaches are compared against the LRPCA baseline released as part of the PaSC distribution [2]. Prior to this competition, the LRPCA baseline represented the best performance seen by the organizers on the PaSC data by a non-commercial algorithm.

4.1. Advanced Digital Science Center

For face detection and alignment, the ADSC group used the eye-coordinates provided by the competition organizers. For each still image and image frame in a video clip, the algorithm cropped, scaled and resampled the face so as to create a 64 by 64 face chip with eye centers at a fixed location and a constant 28 pixels between the eyes.

ADSC's approach treated each face video as an unordered set of image instances. For both video frames and still frames, their algorithm extracted block-wise local binary patterns (LBP) [6] and scale-invariant feature transform (SIFT) [12] features using 8×8 and 4×4 blocks, yielding 3776-dimensional and 2048-dimensional features, respectively. A whitened principle component analysis (WPCA) transformation is then applied within each feature space, to reduce the feature vector dimensionalities to 700 and 500, respectively.

A final similarity score between a pair of videos or a video and a still image is obtained as follows. First, the cosine similarity between pairs of target and query frames is computed. Then, the side-information linear discriminant

analysis (SILD) [16] method is used to learn a discriminative projection for these scores. Finally, the mean similarity score is used to express similarity between the sets of individual comparisons between frames.

4.2. CPqD

The CPqD approach applies inter-session variability modeling (ISV) to face recognition using Gaussian mixture models (GMM) [15]. Using the eye coordinates provided with the competition, the faces are scaled, rotated and cropped to 64×80 pixels. Afterward, each cropped image is processed using Tan & Triggs normalization [21], in order to reduce the illumination effects. In the feature extraction step, each normalized image is decomposed into blocks of 12x12 pixels that overlap by 11 pixels in each direction, and the pixels of each block are normalized to zero mean and unit variance. For each normalized block, the 45 lowest frequency 2D discrete cosine transform (2D-DCT) coeffcients are extracted [20].

Enrollment using the ISV approach considers a client model to be an offset of two elements: the client offset and the session offset. The first one is estimated by a Maximum *a posteriori* (MAP) adaptation from a Universal Background Model GMM (UBM-GMM) [19]. The intuition is that a client model is a mean shift from a general GMM. The second offset compensates session effects such as, pose, illumination and expression. These session effects are assumed to be contained in a linear subspace (U subspace) of the UBM-GMM mean super-vector. Both elements were trained using the training set of the PaSC database and the frontal face images of the BioCPqD database [22]. More details about this approach can be found in [15].

Due to a tight deadline to comply with the competition deadline, only 2 frames of each query video were used in similarity score generation. It is hoped that the performance of this approach will improve once additional query video content is used. The CPqD team also credits the free signal processing and machine learning toolbox Bob [1] for providing useful feature extraction and score generation support.

4.3. Stevens Institute of Technology

The SIT group approached the video face recognition problem with a compact and effective video face representation, the Eigen-PEP representation [10]. The Eigen-PEP representation is built upon the probabilistic elastic part (PEP) model [8, 9]. By design, the Eigen-PEP representation is a pose-invariant, unified representation for faces in images and videos. In practice, despite the number of frames in a video, its Eigen-PEP representation is a low-dimensional (typically 100 dimensional) vector. Using the Labeled Faces in the Wild (LFW) dataset [5], as training data, each image is transformed into a set of spatial-appearance descriptors. The spatial-appearance descriptor is the local appearance descriptor (e.g. SIFT) augmented by its spatial location in the image. Then, given an image, a dense collection of descriptors are extracted over a 3-scale image pyramid with scaling factor 0.9 in a 8×8 sliding window and 2-pixel spacing. The result is a representation of each image in terms of a set of descriptors. Considering the symmetrical structure of the human face, the descriptors extracted from the mirror-image flipped version of the image are added to the descriptors set.

The PEP-model is next built from all the training spatial-appearance descriptors. The PEP-model is a Gaussian Mixture Model with spherical Gaussian components. With the PEP-model, given a face image as a set of descriptors, each Gaussian component identifies the descriptor which induces the highest probability. The PEP-representation for a face image is constructed by concatenating the appearance parts of the selected descriptors in sequence. Then the PEP-representation is compacted by Principle Component Analysis (PCA), resulting in the final Eigen-PEP representation. In practice, only the 100 principal dimensions with the largest sample variance are retained so that the size of the Eigen-PEP representation for a video is only 400 bytes. Given a face video (or image as one-frame video), its Eigen-PEP representation is the average of the Eigen-PEP representations of all its frames. The face matching step is carried out using a joint Bayesian classifier [3].

In the PaSC challenge, a PEP-model learned from the LFW dataset with 1024 components was used. Then an Eigen-PEP representation for the still images and videos was built. Both still frames and frames from video were cropped to a normalized size: 150×150 pixels with interpupillary distance of 44 pixels. Cropping and normalization was carried out using the eye locations provided by the competition organizers. The similarity scores submitted were the result of finding first the Eigen-PEP representations for a target and query pair and then running the joint Bayesian classifier trained over the LFW dataset.

4.4. University of Ljubljana

The face recognition system used by the University of Ljubljana represents an evolution of the recognition system that achieved the best performance at the face-recognition competition held in conjunction with International Conference on Biometrics 2013 [4], [23]. As shown in Figure 4, the system exploits multiple facial representations, data whitening and the recently introduced probabilistic linear discriminant analysis (PLDA) [11] together with two different scoring techniques.

In the first step, the system uses the eye-coordinates provided by the competition organizers to geometrically normalize each given facial image. Once the image is normalized, the system crops the facial region from the input image

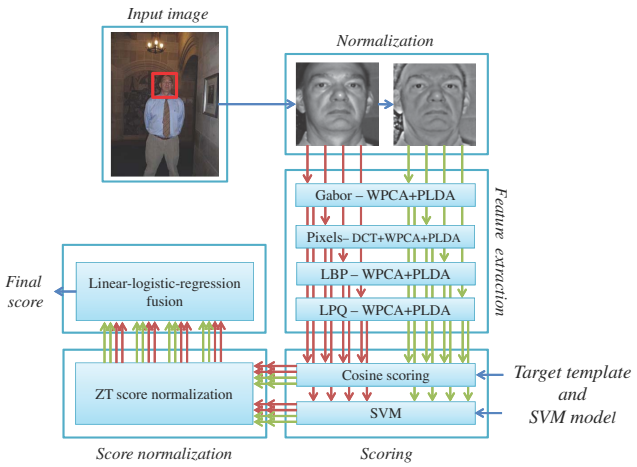

Figure 4. Block diagram of the face recognition system of the University of Ljubljana

and scales it to a fixed size of 50×50 pixels. The scaled facial image is then converted to gray-scale and photometrically normalized using the retina-based normalization technique presented in [24]. This procedure results in two facial images (i.e, a grey-scale and a photometrically normalized one) that form the basis for feature extraction.

In the feature extraction step, four different image descriptors, i.e., Gabor magnitude features, pixel intensities, LBP histograms and local phase quantization (LPQ) pattern histograms, are first computed from the two facial images and then subjected to a PCA-based whitening transform, whose goal is two-fold: dimensionality reduction and data whitening. The dimensionality reduction and whitening step is followed by the computation of PLDA feature vectors for each image descriptor.

In the last step, the computed PLDA feature vectors (2 input images $\times$ 4 descriptors = 8 feature vectors) are subjected to a cosine-based as well as a support vector machine (SVM)-based scoring procedure, which given a target template and target SVM model produces two distinct scores for each feature vector. Hence, a total of 16 scores are calculated in the scoring step. All 16 scores are normalized using a special type of ZT-normalization and ultimately combined using a linear-logistic-regression-based (LLR) fusion technique to produce the final score for the given verification trail.

It should also be noted that all parts of the UNILJ system were trained exclusively on the (video) training data of the PaSC database.

5. Results

The evaluation protocol used in this competition is in keeping with prior face recognition evaluations [17] and adheres to best practices for biometric systems [14]. In particular, algorithms are responsible for returning similarity

Table 3. Verification rates at FAR=0.01 for the video-to-video (Exp. 1) and video-to-still (Exp.2) tasks.

Group	Algorithm	Exp. 1	Exp. 2
ADSC	LPB-SIFT-WPCA-SILD	0.09	0.23
CPqD	ISV-GMM	0.05	0.11
SIT	Eigen-PEP	0.26	0.24
Ljub	PLDA-WPCA-LLR	0.19	0.26
CSU	*LRPCA Baseline*	*0.08*	*0.10*

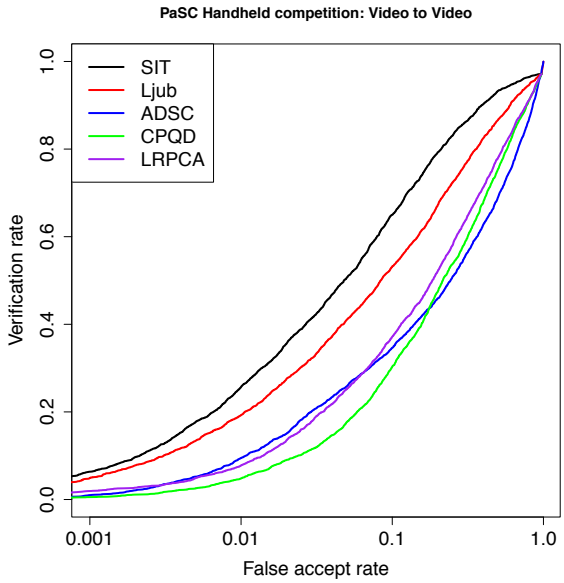

Figure 5. ROC for the Video-to-video Experiment.

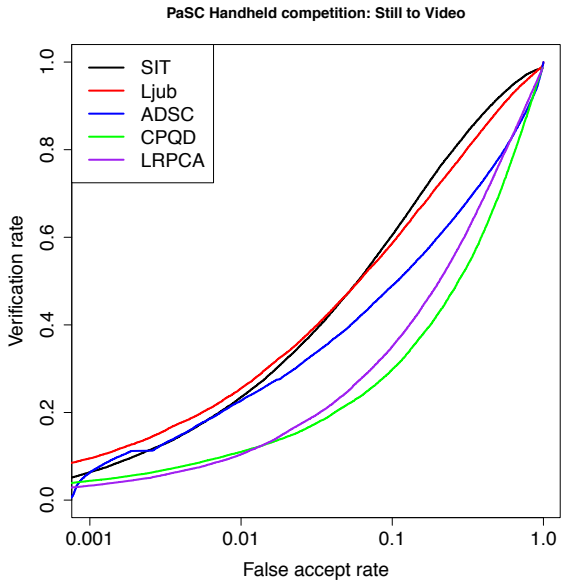

Figure 6. ROC for the Still-to-video Experiment.

scores for pairs of biometric signatures. In the Video-to-video Experiment both the target and query signatures are videos. In the Still-to-video Experiment the query signatures are videos and target signatures are still frame images. Recognition performance is characterized in terms of a verification task, namely the task of determining if a pair of signatures are of the same person. This formulation gives rise to two classes of errors: false matches and false rejects. In biometrics, it is arguably never the case that these are of equal importance. Generally, systems are expected to operate at low false accept rates. Consequently, the single performance number chosen to characterize the performance of an algorithm on an experiment is the verification rate (true positive rate) at a false accept rate (FAR) of 0.01. Table 3 summarizes the results for the competition. The corresponding ROC curves are presented in Figures 5 and 6. On the Still-to-video Experiment, matching a video to still images, all participants demonstrated performance above that of the LRPCA baseline. On the Video-to-video Experiment, matching videos to videos, three participants demonstrated performance above the baseline. Two participants in particular, SIT and Ljub demonstrated performance well above the baseline.

These results were checked for statistical significance using McNemar's test applied to face verification performance [25]. At a p-value threshold of 0.05 these differences in performance are all statistically significant. Two cases from the Still-to-video Experiment are close enough to fail this same test at a threshold of 0.01. The p-value for the comparison between ADSC and SIT is 0.024. The p-value for CPQD and LRPCA is 0.033.

6. Covariate Analysis

The performance results on the Video-to-video Experiment for the two top performing algorithms, Ljub and SIT, have been further analyzed with respect to a set of covariates in order to better understand what factors contribute most to differences in performance. The methodology used in this analysis is more fully explained in [7]. Briefly, the analysis expresses expected verification performance associated with different covariate settings, for example for men versus women.

Figure 7 summarizes the results of this analysis showing main effects plots for nine factors. The x-axis is the factor name with its discretized level settings and the y-axis is the verification rate at FAR=0.01. A large magnitude indicates that a factor has a greater effect on the performance while a small magnitude indicates that the factor has a lesser effect. The analysis is broken into three parts.

The Image/Video Group examines the influence of the range of yaw, roll, and size of a face, and face detection confidence in a video. All four covariates are measured for each frame by the PittPatt SDK 5.2.2. Real-valued covariates for a video sequence are computed by averaging over all frames where an estimate could be computed. For representation in Figure 7, the video covariates were subsequently divided into three levels: S (small), M (middle), L (large). Thus, for example, the verification rate for Size L is notably higher for both algorithms compared to S or M, and indicates a not surprising tendency for both algorithms to perform better when on average faces appear larger in the videos. Perhaps the most surprising aspect of the results for the Image/Video Group is how often these covariates show only very small influence over verification performance. In particular, the confidence covariate is striking because one might well expect better verification performance when faces are found with greater confidence, and this is not evident at all for the Ljub algorithm. A modest effect showing better verification for videos with higher face finding confidence is seen for the SIT algorithm.

The Environment Group examines the influence of where videos were collected, environment, and camera/sensor. A total of 6 locations and 5 handheld video cameras were used to collect the Video-to-video Experiment data. There is a confounding of location (Env.) and the actions (Act.) because data was collected by a script in which generally one action was specified for one data collection location. Consequently, in this analysis, the two are treated as a single covariate. For this group, there are more combinations than are being presented. For example, there are a total of 16 possible pairings of the 6 locations. Three are chosen to emphasize the range of influence. Specifically, the combinations with the lowest, median and highest verification rates. The pairs of activities for the lowest category may be described as people recorded outdoors under a canopy. The median combination compares videos taken in an office to videos taken under the canopy. The highest combination compares people indoors approaching an easel to people picking up a phone. A similar strategy is used to select three combinations of sensors. This covariate is confounded with environment because most cameras were associated with a single environment. The most striking finding from this Environment Group is the magnitude of the effect. While the following is speculation, it seems fair to conclude that the hidden factors related to locations, actions and sensors are more importance than any other factors in this study.

The Subject Group examines the influence of relative difficulty between 'easy' and 'hard' people along with gender and race. The subjectID settings were divided into the three levels (S, M, L) based on the ordered individual performance. In other words, the 265 people were divided into three roughly equal sized groups based upon verification performance. So, for example, the S group consists of people with the lowest verification performance: the 'hard' people. The gender factor has two levels (M: male, F: fe-

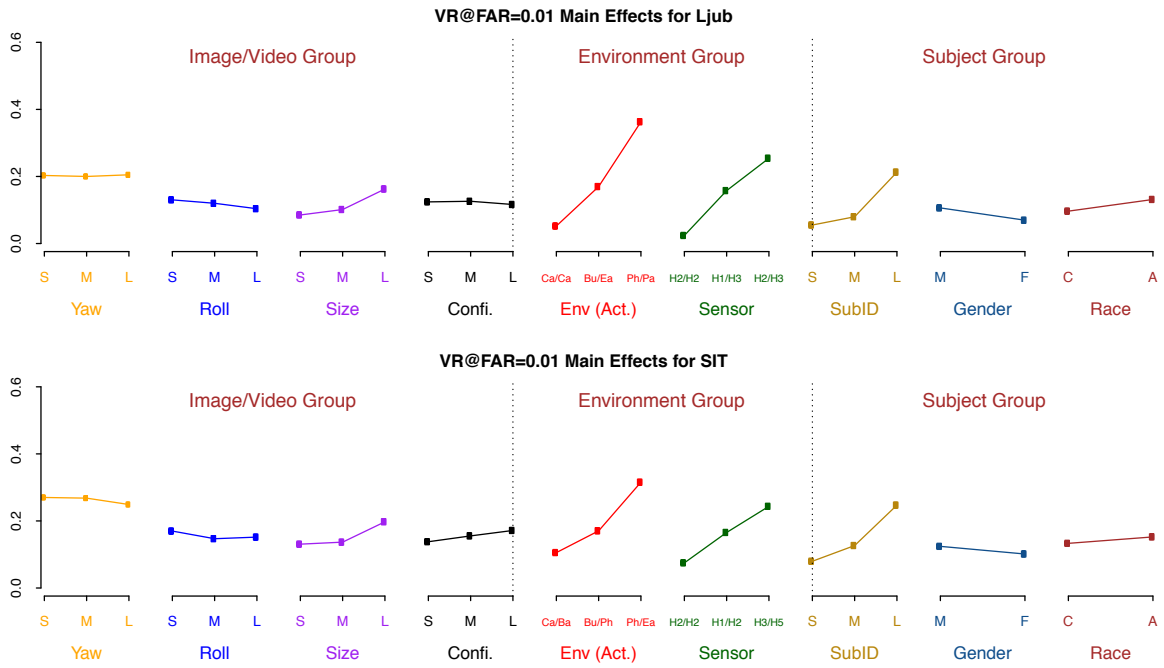

Figure 7. Verification performance broken out by different factors for the Ljub and SIT algorithms on video-to-video.

male), and the race with the two levels (C: Caucasian, A: Asian). The subject effect here is large, the third largest by overall change in verification rate. This is consistent with prior studies showing that performance varies considerably between people. However, it is important to recognize this analysis simply shows a significant difference when people are grouped after the fact by difficulty. The analysis does not address what is making performance vary by person or even if the difference is actually intrinsic to the people themselves. The gender and race effects are more modest and in both cases are consistent with previous reported results [13].

7. Conclusion

The PaSC Video Face and Person Recognition Competition was organized using an existing face recognition challenge data set of significant difficulty. Both experiments highlighted face recognition using video from consumer grade handheld video cameras. The competition results support the perception that person recognition in this context is difficult: verification rates at a 0.01 false accept rate were no higher than 0.26 for any of the participant's algorithms. The previous best performance achieved by CSU with the LRPCA baseline served as a good point of departure for the competition, highlighting the collective improvement achieved by the competition participants using more capable techniques.

In this competition, the best algorithms nearly tripled verification rates relative to the baseline. This is an excellent start, particularly in light of the fact that participants worked under very tight time constraints. At the same time, there is considerable 'headroom' in the handheld video portion of PaSC. As a new instance of the challenge problem model (that originated in biometrics with FERET over twenty years ago [18]) we hope that PaSC will feed continued focused research on a difficult problem and help the research community to chart its progress as it moves forward with newer and more capable algorithms. Toward that end, a Video Person Recognition Evaluation is being organized as part of FG 2015 [7] in order to provide an additional opportunity for groups to work on and document progress on the PaSC video data.

References

[1] A. Anjos, L. El-Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel. Bob: A free signal processing and machine learning toolbox for researchers. In *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, pages 1449–1452, New York, NY, USA, 2012. ACM.

[2] J. Beveridge, P. Phillips, D. Bolme, B. Draper, G. Givens, Y. M. Lui, M. Teli, H. Zhang, W. Scruggs, K. Bowyer, P. Flynn, and S. Cheng. The challenge of face recognition from digital point-and-shoot cameras. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pages 1–8, Sept 2013.

[7] http://www.cs.colostate.edu/~vision/pasc/fg2015

[3] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun. Bayesian face revisited: A joint formulation. In *Computer Vision–ECCV 2012*, pages 566–579. Springer, 2012.

[4] M. Günther, A. Costa-Pazo, C. Ding, E. Boutellaa, G. Chiachia, H. Zhang, M. de Assis Angeloni, V. Struc, E. Khoury, E. Vazquez-Fernandez, D. Tao, M. Bengherabi, D. Cox, S. Kiranyaz, T. de Freitas Pereira, J. Zganec-Gros, E. Argones-Rúa, N. Pinto, M. Gabbouj, F. Simões, S. Dobrisek, D. González-Jiménez, A. Rocha, M. Uliani Neto, N. Pavesic, A. Falcão, R. Violato, and S. Marcel. The 2013 face recognition evaluation in mobile environment. In *The 6th IAPR International Conference on Biometrics*, June 2013.

[5] G. B. Huang, M. J. Jones, and E. Learned-miller. Lfw results using a combined nowak plus merl recognizer. In *In ECCV faces in real-life images workshop*, 2008.

[6] J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth. 3d assisted face recognition: A survey of 3d imaging, modelling and recognition approachest. In *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pages 114–114, June 2005.

[7] Y. Lee, P. J. Phillips, J. J. Filliben, J. R. Beveridge, and H. Zhang. Generalizing face quality and factor measures to video. In *Proceedings of the 2014 International Joint Conference on Biometrics*, September 2014.

[8] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic matching for pose variant face verification. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3499–3506. IEEE, 2013.

[9] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic part model for unsupervised face detector adaptation. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 793–800, Dec 2013.

[10] H. Li, G. Hua, X. Shen, Z. Lin, and J. Brandt. Eigen-Pep for Video Face Recognition. Submitted to ECCV and under review., 2104.

[11] P. Li, Y. Fu, U. Mohammed, J. Elder, and S. J. Prince. Probablistic models for inference about identity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):144–157, 2012.

[12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, Nov. 2004.

[13] Y. M. Lui, D. Bolme, B. Draper, J. Beveridge, G. Givens, and P. Phillips. A meta-analysis of face recognition covariates. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*, pages 1–8, Sept 2009.

[14] T. Mansfield, G. Kelly, D. Chandler, and J. Kane. Biometric product testing final report. *Contract*, 92(4009):309, 2001.

[15] C. McCool, R. Wallace, M. McLaren, L. El Shafey, and S. Marcel. Session variability modelling for face authentication. *Biometrics, IET*, 2(3):117–129, September 2013.

[16] Meina Kan, Shiguang Shan, Dong Xu and Xilin Chen. Side-Information based Linear Discriminant Analysis for Face Recognition. In *Proceedings of the British Machine Vision Conference*, pages 125.1–125.0. BMVA Press, 2011.

[17] 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 *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*, volume 1, pages 947–954. IEEE, 2005.

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

[19] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10(1–3):19 – 41, 2000.

[20] C. Sanderson and K. K. Paliwal. Fast features for face authentication under illumination direction changes. *Pattern Recogn. Lett.*, 24(14):2409–2419, Oct. 2003.

[21] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Image Processing, IEEE Transactions on*, 19(6):1635–1650, June 2010.

[22] Violato, M. U. Neto, F. Simões, T. de Freitas Pereira, and M. Angeloni. Biocpqd: uma base de dados biométricos com amostras de face e voz de indivíduos brasileiros. In *Cadernos CPqD Tecnologia*, volume 9, 2013.

[23] V. Štruc, J. Žganec Gros, S. Dobrišek, and N. Pavešić. Exploiting representation plurality for robust and efficient face recognition. In *Proceedings of the 22nd Intenational Electrotechnical and Computer Science Conference (ERK'13)*, pages 121–124, Portorož, Slovenia, September 2013.

[24] N. Vu and A. Caplier. Illumination-robust face recognition using retina modeling. In *Proceedings of the the International Conference on Image Processing (ICIP'09)*, pages 2335–2338, 2009.

[25] W. Yambor, B. A. Draper, and J. R. Beveridge. Analyzing pca-based face recognition algorithms: Eigenvector selection and distance measures. In H. Christensen and J. Phillip, editors, *Empirical Evaluation Methods in Computer Vision*. World Scientific Press, Singapore, 2002.