Frontal to Profile Face Verification in the Wild

Soumyadip Sengupta[1], Jun-Cheng Chen[1], Carlos Castillo[1], Vishal M. Patel[2], Rama Chellappa[1], and David W. Jacobs[1]

[1]Center for Automation Research, University of Maryland, College Park, MD 20740, USA.
[2]Department of Electrical and Computer Engineering Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA.

Abstract

We have collected a new face data set that will facilitate research in the problem of frontal to profile face verification 'in the wild'. The aim of this data set is to isolate the factor of pose variation in terms of extreme poses like profile, where many features are occluded, along with other 'in the wild' variations. We call this data set the Celebrities in Frontal-Profile (CFP) data set. We find that human performance on Frontal-Profile verification in this data set is only slightly worse (94.57% accuracy) than that on Frontal-Frontal verification (96.24% accuracy). However we evaluated many state-of-the-art algorithms, including Fisher Vector, Sub-SML and a Deep learning algorithm. We observe that all of them degrade more than 10% from Frontal-Frontal to Frontal-Profile verification. The Deep learning implementation, which performs comparable to humans on Frontal-Frontal, performs significantly worse (84.91% accuracy) on Frontal-Profile. This suggests that there is a gap between human performance and automatic face recognition methods for large pose variation in unconstrained images.

1. Introduction

Face recognition for unconstrained images is a challenging problem, due to variation in pose, illumination, expression, age and occlusion. A significant challenge of pose variation occurs when features from the whole face are not visible. These situations appear often in many real world scenarios like, surveillance and photo-tagging, where it is quite natural for a person not to face the camera. In this paper we plan to study the effect of pose variation, isolated as a factor, in the presence of other 'in the wild' variations. One such case of interest is matching a frontal face facing a camera to a profile face facing away from the camera. The features available in both these views vary significantly and are therefore difficult to match.

We define a 'near profile' pose as one that obscures many features; specifically the second eye. This roughly corresponds to yaw greater than 60 degrees. We define 'near frontal' as those cases where both sides of the face are almost equally visible and the yaw is within 10 degrees of purely frontal. The main motivation of this work is to study face recognition in the presence of such extreme pose variation 'in the wild'.

Face recognition has changed significantly over the past decade. Starting with constrained, carefully acquired images, the community has turned its attention to the distinct problem of face recognition in unconstrained settings. There are many data sets that have aided this progress. Labeled Faces in the Wild (LFW) [17] was acquired to study the problem of face recognition from unconstrained images and consists of images collected via the Internet. The Multipie [12] data set encourages face recognition in the presence of pose, illumination and expression variation in a controlled environment. One of the shortcomings of the LFW data set is that it doesn't offer a high degree of variation in terms of pose, like the variation in pose present in Multipie. Large pose variation has been shown to be a major challenge in face recognition. In this paper we propose a new data set, which in principle is a mixture of constrained and unconstrained settings. We collect images from the internet, which are unconstrained, but filter them out to match specific 'frontal and 'profile' poses. This allows us to study the problem of pose variation in a more controlled way while all other variations are unconstrained. We call this data set 'Celebrities in Frontal-Profile data set' (CFP). We believe solving this problem with extreme pose variation will bring more success to the general problem of unconstrained pose variation, especially in cases of surveillance and photo-tagging.

The data set contains 10 frontal and 4 profile images of

Figure 1. Sample Images from proposed Celebrities in Frontal-Profile (CFP) data set

500 individuals. Similar to LFW, we have defined 10 splits, each containing 350 same and 350 not-same pairs. The task is face verification.

To understand how difficult the problem of frontal to profile comparison in the wild is compared to frontal to frontal unconstrained image recognition, as in LFW, we form two separate experiments of Frontal-Profile and Frontal-Frontal face verification. We evaluated some state-of-the-art face recognition algorithms and human responses on this experiment to obtain a sense of the difficulty posed by this new data set. We found that humans achieve 94.57% accuracy on Frontal-Profile as compared to 96.24% on Frontal-Frontal. On the other hand we observe that for most state-of-the-art algorithms, verification accuracy degrades at least by 10% from Frontal-Frontal to Frontal-Profile, meaning that the error rate more than doubles. We notice that among different types of hand-crafted features Fisher Vector [31] performs better than HoG [10] and LBP [1]. In restricted settings, Fisher Vector with a metric learning, SubSML [5], achieves 80.63% accuracy on Frontal-Profile and 91.3% accuracy on Frontal-Frontal. These findings show that face recognition with large pose variation in an uncontrolled environment is still an open research problem. Our data set attempts to enable research in this important problem to achieve or surpass human accuracy on this task.

Learning features via a deep neural network, has brought huge success in unconstrained face recognition. Many of the current state-of-the-art deep learning implementations [33], [32], [27], are not publicly available. We used a deep learning algorithm [9] which achieved near human accuracy (96.4%) on Frontal-Frontal face verification. However even this algorithm falls short by 11% in accuracy (84.91%) on Frontal-Profile verification as compared to humans on this task. This network was trained with approximately 400K unconstrained images of around 10K subjects. Certainly one can try to train a network on profile images alone. However collecting millions of profile images, which most state-of-the-art deep networks require for training, is difficult. For example, if we search 'Google Image' for profile images of a person, less than 2% of top 100 images are actually profile. This means a huge amount on post-processing needs to be done to remove wrong poses and identities to collect millions of images. So can we do something better without trying to collect a huge number of profile images? This requires an advance in research in face recognition with large pose variation and our work is aimed to provide a benchmark for these future researches.

Most algorithms require some facial key-points either for aligning [33] or extracting features [5]. Current state of the art key-point detectors fail in the case of near profile poses. Along with the images, our data set provides hand-annotated key-points for profile faces to encourage research in automatic key-point detection for profile faces.

To summarize, frontal to profile face recognition in the wild is important because:

- It occurs commonly in many applications.
- The performance of existing algorithms degrades significantly when comparing frontal faces to profile faces in the wild.
- The performance of humans in frontal to profile face comparisons is only slightly worse compared to frontal to frontal.

In Section 2 we discuss some of the related work in automatic key-point extraction, face recognition across pose variation, recognition in real world images and some current data sets. Section 3 presents detailed discussion of the proposed CFP data set. Section 4 discusses the algorithms, whose performance are compared on our proposed CFP data set, followed by their performance evaluation.

2. Related Work

The vast majority of face recognition methods that attempt to handle pose variation use key-points [5], [7], [33]. Key-points are used to align the images, extract features at specific locations and warp faces to a canonical view. There has been quite a bit of progress in automatic key-point detection and there are many publicly available key-point detection algorithms [3], [38] and [36]. State of the art methods perform very well on frontal images, but the accuracy of key-point detectors degrades as the yaw or pitch angle of the face increases. Motivated by the above observation, we include dense ground truth key-points with our data set, so that face recognition across pose can improve while key-point detectors get to the point where they can handle the full spectrum of pose variation. Also these manually annotated key-points serve as a benchmark for future researchers trying to develop automatic key-point extraction algorithms for profile faces.

In addition to facial key-point detection, in this section we will discuss two streams of work: (1) face recognition with pose variation and (2) face recognition 'in the wild'. The first line of work in general depends on carefully acquired constrained image data sets that focus on pose, illumination and expression variations. Some important data sets along these lines are Yale-B [11], FERET [24], CMU PIE [30] and Multipie [12]. The second line of work is developed around data sets which contain unconstrained images. Labeled Faces in the Wild (LFW) [17] has become the de facto benchmark for face recognition in unconstrained settings. Average results on this data set have increased from 70% to 99% in the past 8 years. The LFW data set is based on a face verification protocol, in restricted and unrestricted settings. In the restricted setting, one is only allowed to use pair-wise information provided in the splits. However, in unrestricted setting, one can use identity information to build additional pairs not explicitly listed in the training data or use outside training data. Some other data sets which also provide unconstrained images are PubFig [19] and the YouTube Face data set (YTF) [35].

Face Recognition across pose : Researchers have used several different approaches to solve the problem of face recognition with pose variation. One such popular technique is to fit a Morphable model to the face and warp it to some canonical view. This line of work started in [4] and exploded as a general model fitting technique; for example Generic Elastic Models (GEM) [25] and Active Appearance based Models for Pose normalization [2]. However these methods tend to work well for small degrees of pose variation and for faces without 'in the wild' variation. Another type of methods are based on subspace learning. These methods mostly use Canonical Correlation Analysis (CCA) [14] or Partial Least Square (PLS) [29]. Recently [20] and [28] (27.1% identification accuracy for Frontal-Profile in Multipie) have shown good results over the Multipie and CMU PIE data sets, considering an identification protocol rather than verification. However it has not been demonstrated that these methods will actually work for 'in the wild' images. Another direction of pose invariant research is to develop a method to directly compare faces in two poses using stereo matching [6], which performs well short of state-of-the-art on real world images such as those in the LFW data set. Another line of pose invariant face recognition research deals with generative models. These methods assume that there is a latent factor that produces different identities and different poses are generated by another latent variable. Recently [26] have shown good performance in constrained data sets such as FERET [24]. Along the same line [22] produced good results on unconstrained data set like LFW (90.07% verification accuracy in unrestricted setting). Attribute based recognition [19] is another approach to the problem, which is also potentially invariant to pose variation, although it is not clear that attributes can be obtained as accurately on profile faces as on frontal faces. Most of these method depend on good alignment across pose, which is hard to obtain for our proposed CFP data set.

Face Recognition on unconstrained images : In this section we discuss those methods that have produced good performance over the LFW data set and other 'in the wild' data sets. One general technique adopted by many researchers is to develop metric learning approaches that can learn a transformation of the feature space to reduce the variability, which is important for unconstrained images. Cosine Similarity metric learning [23] and Similarity metric learning [5] (86.73% in unrestricted setting) have produced good results on the LFW data set. Researchers have also developed other metric learning approaches [13], [18] along with Deep metric learning approaches [15] [16]. The Joint Bayesian model [7] (90.90% accuracy in unrestricted setting) and [8] (93.18% accuracy) performs well on LFW. However these methods generally need identity information during training and thus can only be used with the unrestricted protocol (where one can use identity information or outside training data).

Researchers have also concentrated on feature extraction techniques other than traditional SIFT, LBP or HoG to provide a higher level representation. One such efficient method uses Fisher Vector encoding [31] (87.47% in restricted settings) and [21] (84.08% in restricted settings). However they are not robust against large pose variation. Researchers have moved from hand crafted features to trained features using Deep networks namely CNNs (Convolutional Neural Networks). Some successful applications are Deepface [33] (97.35%), DeepID [32] (99.47%) and FaceNet [27] (99.63%), which have shown to be the current state of the art among algorithms on LFW in the unrestricted setting with outside training data. Since most of these algorithms are not publicly available, we used a different deep learning technique [9] for extracting features. We show that it performs as humans on Frontal-Frontal but falls short by a large margin on Frontal-Profile compared to humans. However this network is trained on unconstrained images. One can certainly try to train a network on a large amount of profile images, which is very difficult to collect. It will be also interesting to observe how researchers come up with new methods that can tackle pose variation without explicitly collecting millions of profile images.

3. Celebrities in Frontal-Profile data set

We have collected a data set of unconstrained faces in both frontal and profile poses. The experimental protocol is built upon face verification. Unlike LFW, we decided to balance the data set by choosing only a fixed number of frontal and profile images per individual. We will make this data set publicly available.

3.1. Collection Set-up

To collect our data set we started by generating a list of individuals. We decided to maintain balance in the data set by choosing almost equal numbers of males and females and maintaining as much racial diversity as possible. We chose a roughly balanced set of politicians, athletes and entertainers. In order to collect frontal and profile images we downloaded hundreds of images of each individual for frontal and profile respectively. To search for profile images, we used keywords 'profile face' and 'side view'. Though most of the frontal images are correct in terms of pose or identity, there were lots of non-profile images in the downloaded profile images. Next we cleaned up the data set by deleting incorrect identities and poses by using Amazon Mechanical Turk. We defined 'frontal' pose as those images where both sides of the face are almost the same area of the image and 'profile' pose as those images where one eye is completely visible and less than half of the second eye is visible. Roughly, these definitions mean : within 10 degrees yaw variation for 'frontal' and more than 60 degrees yaw variation for 'profile'. To make this technical criteria clear we provided example images to the Amazon Mechanical Turk workers for reference. We also ran a face detector [38] to further verify that these images are indeed faces and they satisfy the yaw variation constraints. We have a total of 500 individuals and we choose to keep 10 frontal and 4 profile images per person in the data set.

We crop the frontal images by running a face detector [3]. However these state-of-the-art face detectors perform imperfectly on our profile images. Then we set up an Amazon Mechanical Turk job where we asked workers to manually crop the faces from all the profile image. For frontal images we extract key-points by running a facial key-point detector [3]. However none of the state-of-the-art detectors, that we tried [38], [3], work well for profile faces. We acquired labeled key-points from workers via Amazon Mechanical Turk. We present the workers with examples of manually clicked profile key-points and ask them to mark the same. Since in profile only one side of the face is visible, we choose to keep key-points only on one side for frontal faces also. This is to ensure that we always have correspondence in key-points across all images. An example of key-points in frontal and profile is shown in Figure 2. Based on the key-points both the frontal and profile images are cropped. We create a tight bounding box on the face based on the key-points and enlarge it consistently to accommodate large regions of the face.

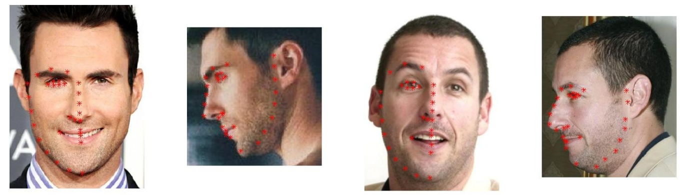


Figure 2. key-points for Frontal-Profile Images

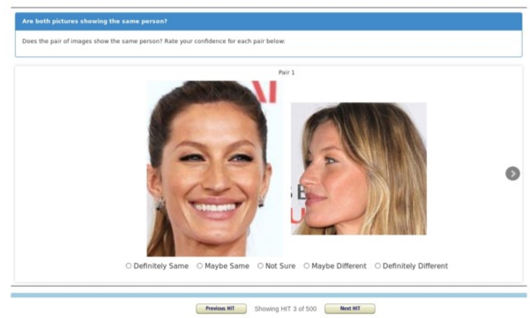


Figure 3. Amazon Mechanical Turk window for obtaining human performance on CFP data set

3.2. Experimental Protocol

We divide the data into 10 splits with a pairwise disjoint set of individuals in each split. For each split we have 50 individuals. We randomly generate 7 same and 7 not-same pairs for each individual, thus producing 350 same and 350 not-same pairs per split. This is done both for Frontal-Frontal and Frontal-Profile experiments. In the end, we have 7000 pairs of faces for both Frontal-Frontal and Frontal-Profile experiment. Note that our choice of generating pairs in the split is balanced in terms of number of pairs

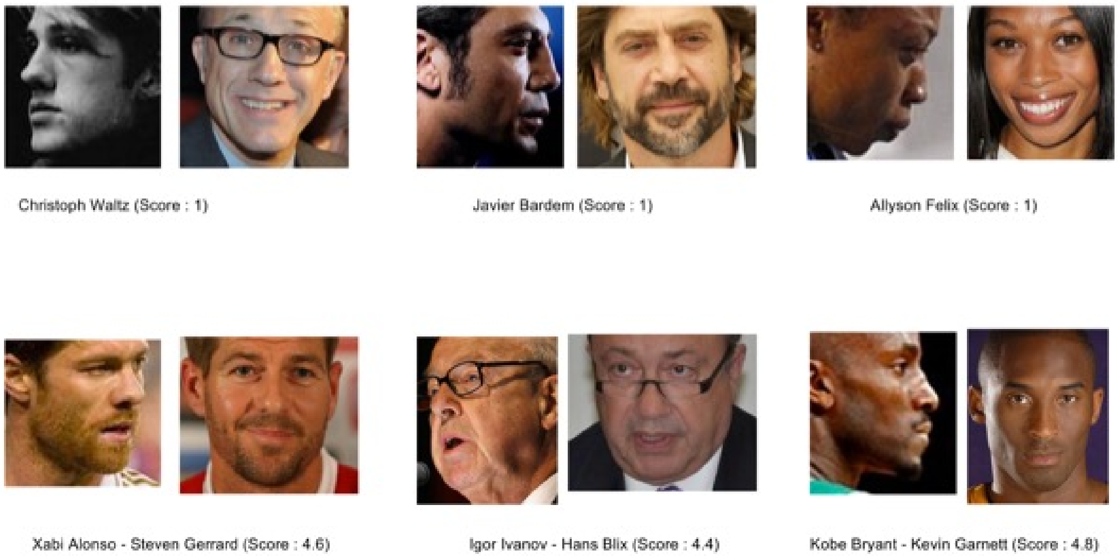


Figure 4. Human performance on CFP data set. (Top row) Top 3 mistakes on same-pairs. (Bottom row) Top three mistakes on not-same pairs. Higher score means more similar in a scale of 1-5.

per person, unlike LFW and YTF data set. Like LFW, the protocol is to test on one split while training on the remaining nine. Unlike LFW, we do not have separate View 1 and View 2 (two views are provided in LFW to develop models and validate on one view and finally test on another), due to a lack of data (as collecting profile images is more difficult), therefore similar to YTF we only have one view. For each split the evaluator is allowed to choose the parameter of the classifier via cross-validation over the training data set only. We report the average Accuracy, Equal Error Rate (EER), Area Under the Curve (AUC) and the ROC curve.

Similar to LFW, there can be a 'restricted' setting, where the evaluator is only allowed to use same-different pairs in training the classifier and 'unrestricted' setting, where the evaluator can use identity labels of the training images. One can have two more variations depending on the use of outside training images which is of special interest in 'unrestricted' settings. In our experiments we used a 'restricted no outside image' protocol for all algorithms except the deep learning implementation. Deep learning implementation uses 'outside training image' to train the network, but the cosine-similarity metric only uses the same-different pairs of the CFP data set.

3.3. Human Performance

Once the data set is ready we ask the next question, how good are humans on this task? In [19], the authors evaluated human performance on the LFW data set. Humans performed 97.53% on cropped images of LFW. We expected Frontal-Profile verification to be harder than LFW evaluation. With similar cropped images we note that humans perform 94.57% on the frontal-profile experiment in comparison to 96.24% on frontal-frontal experiment of our CFP data set. We note that the degradation of human performance from LFW to the CFP frontal-profile data set is significantly less than what state-of-the-art algorithms show. This also indicates the need for research to tackle large pose variation.

To provide a sense of how hard the problem is for humans, we show in Figure 4 the top mistakes made by humans on Frontal-Profile. Each Mechanical Turk worker scores each image on a scale of 1-5, where higher score means more similar. We show top three mistakes from same pair and not-same pair in the image.

Human experiments are performed via Amazon Mechanical Turk. We show each pair of images to 5 workers who may or may not be familiar with the individual being depicted. We then ask each of them to rate the similarity between pairs on a scale 1 to 5 (where 1 is definitely different and 5 is definitely same). This shows the confidence of the decision from the user. We remove any outlier or erroneous workers and average out their scores to produce the final accuracy and ROC curve. The interface for human evaluation is very simple as depicted in figure 3.

4. Experimental Evaluation

To show the difficulty posed by Frontal-Profile verification in our CFP data set, we evaluate numerous state-of-the-art algorithms on this data set. We look at those algorithms that perform well on unconstrained or 'in wild' data sets like LFW and whose implementations are publicly available. We consider different types of feature extraction techniques like HoG [10], LBP [1] and Fisher Vector [31] along with different metric learning techniques like Sub-SML [5] and others as reported in [31]. Sub-SML [5] appears to be very successful metric learning technique compared to others on the LFW data set. We run the experiment on both Frontal-Frontal and Frontal-Profile in 'restricted' settings. We also used a Deep learning implementation [9] which uses outside images for training the network.

4.1. Algorithms

We use three different types of feature extraction techniques, the details of which are discussed below :

- **HoG** : We extract square patches of width 10, 15, 30, 50 pixels centered around each of the 30 facial key-points. Then we extract HoG features of cell-size 8 from these patches and concatenate them to form a 53k dimensional HoG feature of the face. Multiple-scale patches are used to provide a multi-resolution view of the face. We use the VLFeat [34] implementation of HoG.

- **LBP** : Similar to HoG we extract square patches of size 10, 15, 30, 50 and 100 pixels centered around 30 key-points. We then extracted uniform LBP features (sampling points 16) of radius 1 and concatenate them to form a 36k dimensional LBP feature of the face.

- **Fisher Vector** : We used publicly available code of Fisher Vector and followed the same principle of [31]. However we didn't use horizontal flipping of images to make it consistent with other features. Fisher vector encoding with 512 cluster centers result in a 67,584 dimensional feature.

- **Deep features** : We use the trained network reported in [9]. The authors use a deep network with 10 convolution layers, 5 pooling layers and 1 fully connected layer. The receptive field of the CNN is $100 \times 100 \times 1$. The authors claim that a deeper network with a smaller number of filters is easier to train because it uses fewer parameters and performs better due to high amount of non-linearity. The network is trained on the CASIA-Webface data set [37] with 494,414 images of 10,575 subjects. We only used the network to extract features of dimension 320. We used a simple Cosine similarity measure over this feature.

We use different types of classifiers, mainly based on metric learning. We use publicly available code of Sub-SML [5]. All the features are reduced by PCA to 300 dimension, whitened and then used with Sub-SML. Other than Sub-SML we use Diagonal metric learning (DML) as reported in [31]. We should point out the differences between these different techniques. Diagonal Metric Learning (DML) is learning to weight different feature dimensions, which can be implemented via a linear SVM formulation. Sub-SML learns both a distance metric along with a similarity kernel, and includes a regularization in the formulation, which penalizes too much distortion of these matrices and is implemented via a fast first order method.

4.2. Results

We present the mean and standard deviation of Accuracy, Equal Error Rate (EER) and Area Under Curve (AUC) over the 10 fold experiments for both Frontal-Profile and Frontal-Frontal Experiment in Table 1 and also present the average ROC curves for them in Figure 5.

4.3. Discussion

From the experimental results we can observe the significant drop in performance of all the algorithms from Frontal-Frontal to Frontal-Profile. On the other hand, human performance only deteriorates around 2% from Frontal-Frontal to Frontal-Profile. However most of the algorithms degrade around 10%. For Frontal-Frontal, Deep features produce near-human accuracy. Whereas for Frontal-Profile it falls short of human performance by 11%. This means even though the problem is hard for humans, they perform well compared to current state-of-the-art algorithms. Thus there is a huge room for improvement.

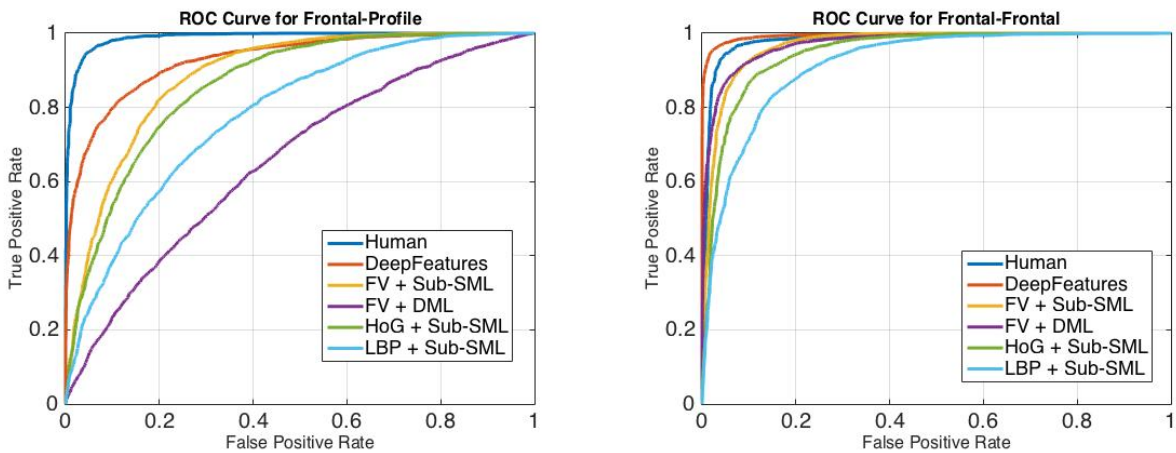
In the restricted protocol, we can see that Fisher Vector with Sub-SML performs best of all the algorithms. We can observe that Fisher vector is a much better feature than HoG and LBP as it is more robust to pose variation. Also HoG and LBP features used in the baseline, need dense sets of 30 facial key-point to extract patches, whereas Fisher Vector only needs 3 key-points for rough alignment. We also compare different metric learning algorithms with Fisher vector as the feature. We note that Sub-SML is much better as a metric learning method due to the regularization used in the formulation. Diagonal Metric Learning (DML) formulation performs significantly worse in Frontal-Profile than in Frontal-Frontal.

We are not able to test state-of-the-art deep learning techniques on LFW, on our CFP data set since they are not publicly available and cannot be replicated with existing public data sets and available resources. We used a deep learning implementation [9], which achieves human accuracy on the Frontal-Frontal data. However it falls far short of human accuracy on Frontal-Profile. It would be interesting to observe the performance of current state-of-the-art deep learning methods on our CFP data set.

In training the deep network [9], the author used 7 key-points on both sides of the face to align the images.

Table 1. Performance comparison on CFP data set (Mean Accuracy and standard deviation over 10 folds)

	Frontal-Profile			Frontal-Frontal		
Algorithm	Accuracy	EER	AUC	Accuracy	EER	AUC
HoG + Sub-SML	77.31 (1.61)	22.20 (1.18)	85.97 (1.03)	88.34 (1.33)	11.45 (1.35)	94.83 (0.80)
LBP + Sub-SML	70.02 (2.14)	29.60 (2.11)	77.98 (1.86)	83.54 (2.40)	16.00 (1.74)	91.70 (1.55)
FV + Sub-SML	80.63 (2.12)	19.28 (1.60)	88.53 (1.58)	91.30 (0.85)	8.85 (0.74)	96.87 (0.39)
FV + DML	58.47 (3.51)	38.54 (1.59)	65.74 (2.02)	91.18 (1.34)	8.62 (1.19)	97.25 (0.60)
Deep features	84.91 (1.82)	14.97 (1.98)	93.00 (1.55)	96.40 (0.69)	3.48 (0.67)	99.43 (0.31)
Human	94.57 (1.10)	5.02 (1.07)	98.92 (0.46)	96.24 (0.67)	5.34 (1.79)	98.19 (1.13)



(a) (b)

Figure 5. Roc curve for (a) Frontal-Profile and (b) Frontal-Frontal

However in profile faces, key-points from both sides are not available. In future we plan to find a way to perform this alignment and fine-tune two separate deep networks on Frontal and Profile images of the data set to improve the performance. There is also further possibility to train metric learning over these features. Sub-SML produced worse result than simple cosine similarity over deep features.

5. Conclusion

This work introduces a new data set which aims to enable the study of face recognition in unconstrained images with large pose variation. We analyzed the performance of several different algorithms using a restricted protocol and showed how all of them degrade from Frontal-Frontal to Frontal-Profile. We also used a deep learning based algorithm and showed that it fails to achieve near-human performance in Frontal-Profile unlike Frontal-Frontal. However there are many alternate ways to improve the trained deep network by separately fine-tuning it over Frontal and Profile images and training additional metric learning approaches over deep features. We plan to address these issues in future and try to develop good deep learning architectures that can handle pose variation without explicitly using millions of Profile images. Our data set also provides this opportunity to other researchers. The gap between current state-of-the-art algorithms on this data set and human performance suggests that there is a lot of room for improvement.

6. Acknowledgments

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2014-14071600012. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

We thank Mr. Wasif Sikder for assistance with collection of the data set.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.

[2] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith. Fully automatic pose-invariant face recognition via 3d pose normalization. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 937–944. IEEE, 2011.

[3] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. Robust discriminative response map fitting with constrained local models. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3444–3451. IEEE, 2013.

[4] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1063–1074, 2003.

[5] Q. Cao, Y. Ying, and P. Li. Similarity metric learning for face recognition. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 2408–2415. IEEE, 2013.

[6] C. D. Castillo and D. W. Jacobs. Using stereo matching with general epipolar geometry for 2d face recognition across pose. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2298–2304, 2009.

[7] 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.

[8] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3025–3032. IEEE, 2013.

[9] J.-C. Chen, V. M. Patel, and R. Chellappa. Unconstrained face verification using deep cnn features. *arXiv preprint arXiv:1508.01722*, 2015.

[10] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[11] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):643–660, 2001.

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

[13] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? metric learning approaches for face identification. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 498–505. IEEE, 2009.

[14] D. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12):2639–2664, 2004.

[15] J. Hu, J. Lu, and Y.-P. Tan. Discriminative deep metric learning for face verification in the wild. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1875–1882. IEEE, 2014.

[16] G. B. Huang, H. Lee, and E. Learned-Miller. Learning hierarchical representations for face verification with convolutional deep belief networks. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2518–2525. IEEE, 2012.

[17] G. B. 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.

[18] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2288–2295. IEEE, 2012.

[19] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and Simile Classifiers for Face Verification. In *IEEE International Conference on Computer Vision (ICCV)*, Oct 2009.

[20] A. Li, S. Shan, X. Chen, and W. Gao. Maximizing intra-individual correlations for face recognition across pose differences. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 605–611. IEEE, 2009.

[21] 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.

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

[23] H. V. Nguyen and L. Bai. Cosine similarity metric learning for face verification. In *Computer Vision–ACCV 2010*, pages 709–720. Springer, 2011.

[24] 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.

[25] U. Prabhu, J. Heo, and M. Savvides. Unconstrained pose-invariant face recognition using 3d generic elastic models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(10):1952–1961, 2011.

[26] S. J. Prince, J. Warrell, J. Elder, and F. M. Felisberti. Tied factor analysis for face recognition across large pose differences. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(6):970–984, 2008.

[27] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. *arXiv preprint arXiv:1503.03832*, 2015.

[28] A. Sharma, M. Al Haj, J. Choi, L. S. Davis, and D. W. Jacobs. Robust pose invariant face recognition using coupled latent space discriminant analysis. *Computer Vision and Image Understanding*, 116(11):1095–1110, 2012.

[29] A. Sharma and D. W. Jacobs. Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 593–600. IEEE, 2011.

[30] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 46–51. IEEE, 2002.

[31] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher Vector Faces in the Wild. In *British Machine Vision Conference*, 2013.

[32] Y. Sun, X. Wang, and X. Tang. Deeply learned face representations are sparse, selective, and robust. *arXiv preprint arXiv:1412.1265*, 2014.

[33] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1701–1708. IEEE, 2014.

[34] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *Proceedings of the international conference on Multimedia*, pages 1469–1472. ACM, 2010.

[35] 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. IEEE, 2011.

[36] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 532–539. IEEE, 2013.

[37] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[38] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2879–2886. IEEE, 2012.