Pose, Illumination and Expression Invariant Pairwise Face-Similarity Measure via Doppelgänger List Comparison

Florian Schroff*, Tali Treibitz*, David Kriegman, Serge Belongie
Department of Computer Science and Engineering
University of California, San Diego
{gschroff,tali,kriegman,sjb}@cs.ucsd.edu

Abstract

Face recognition approaches have traditionally focused on direct comparisons between aligned images, e.g. *using pixel values or local image features. Such comparisons become prohibitively difficult when comparing faces across extreme differences in pose, illumination and expression. The goal of this work is to develop a face-similarity measure that is largely invariant to these differences. We propose a novel data driven method based on the insight that comparing images of faces is most meaningful when they are in comparable imaging conditions. To this end we describe an image of a face by an ordered list of identities from a Library. The order of the list is determined by the similarity of the Library images to the probe image. The lists act as a signature for each face image: similarity between face images is determined via the similarity of the signatures. Here the CMU Multi-PIE database, which includes images of 337 individuals in more than 2000 pose, lighting and illumination combinations, serves as the Library. We show improved performance over state of the art face-similarity measures based on local features, such as FPLBP, especially across large pose variations on FacePix and multi-PIE. On LFW we show improved performance in comparison with measures like SIFT (on fiducials), LBP, FPLBP and Gabor (C1).*

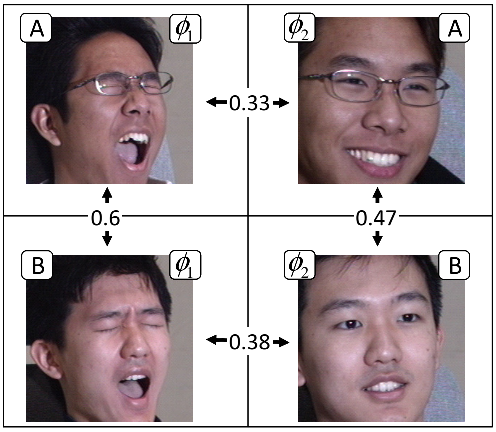

Figure 1. **Face recognition across pose, illumination and expression** is very challenging. Consider person A in two different imaging conditions ϕ_1 and ϕ_2 (top row). Person B is also imaged under these two pose, illumination and expression conditions ϕ_1, ϕ_2 (bottom row). The similarities between the images are calculated using the SSIM measure (Eq. 2), where 1 is the score for identical images and 0 is the lowest score. Images $I(A, \phi_1)$ and $I(B, \phi_1)$ come up as more similar than $I(A, \phi_1)$ and $I(A, \phi_2)$, as anticipated in Eq. (1). This demonstrates the inherent challenge in face similarity across different imaging conditions.

1. Introduction

Face recognition methods have improved dramatically and have a variety of applications including access control, security and surveillance, organization of consumer photo albums, and entertainment. With thousands of published papers on face recognition, progress in research can be chronicled by the representative datasets used for evaluation. Early datasets such as FERET [26], focused research on frontal face recognition under controlled settings akin to an ideal access control application with compliant subjects. With the requirements for security and surveillance that face recognition operate on non-frontal images with uncontrolled lighting and non-compliant subjects, datasets were created with systematic variation in imaging conditions such as pose, lighting, expression, and occlusion in still and video images [11, 13, 22, 29]. For each new dataset, successive methods were devised with increasing performance. Yet, when the methods were applied to the unconstrained photos found in personal photo albums or on the internet, performance was disappointing, and so the latest benchmarks are based on datasets such as Labelled Faces in the Wild (LFW) [12] and Pubfig [20] that are culled from the internet. These datasets not only exhibit variation in pose, lighting, expression and occlusion, but also variations for an individual due to the imaging device, compression, resolution, blur, makeup, hairstyles, eyeglasses,

*The authors assert joint authorship for this work.

2011 IEEE International Conference on Computer Vision 2494
978-1-4577-1102-2/11/$26.00

fatigue and aging as well as greater variations in gender, ethnicity, and age across the subject pool. Yet, these datasets also introduce a bias because of the photos that are exposed on the internet were made public because the photographer/owner found them to be of sufficient quality to be worth posting (internet images are generally much more interesting to look at than surveillance images) and because the faces had to pass certain filters, such a having a detectable face using some available detector.

The lion's share of recognition or verification methods are based on an ability to register pairs of images or an image to a canonical system, often at the pixel level. In purely frontal recognition, this might be accomplished with a similarity transformation whereas for near frontal recognition an affine transformation may be more appropriate. Yet, for a larger degree of pose variation, registration might involve 2-D warps or a 3-D generic/morphable model [5, 7] and the use of fiducial point detection or pose estimation. Registration across expression might use either warps or local patches around fiducial points [11]. Recently, Kumar *et al*. [20] introduced a technique for verification that does not require registration but rather characterizes face images by attributes and similes and compares the attributes and similies for pairs of images rather than registering and comparing the images themselves. Consequently, they have an opportunity to compare images with drastically different pose and expression with partial occlusion, so long as the attribute or simile extraction is invariant to those factors, together with the fiducial detector. Yet to date, this method has only been evaluated on datasets that do not have the extreme imaging conditions (*e.g*., near-profile images and directional lighting) that are found in some of the earlier datasets such as PIE [29].

We introduce a new method for comparing pairs of face images that does not require direct registration of the pair, and in turn allows recognition or verification between images where the mutually visible portion of face is small (*e.g*., from a well lit frontal gallery to a poorly lit, expressive side view). To achieve this, we consider an observation by Moses, Adini and Ullman [23]: "*The variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity.*" Considered formally, let $I(p, \phi)$ denote an image of person p in imaging condition ϕ (pose, illumination, expression, *etc*.) and let $D(I_1, I_2)$ be a measure of distance or similarity of image I_1 and image I_2. An interpretation of this statement is that

$$D[I(A, \phi_1), I(B, \phi_1)] < D[I(A, \phi_1), I(A, \phi_2)] \quad (1)$$

for most image-to-image distance function D such as $\mathcal{L}^2$ or more complex measures, such as FPLBP [32] (Fig. 1 shows a specific case). The quest for imaging condition invariant recognition is to turn this around and find a D' such that

$$\forall \phi_1, \phi_2,\ D'[I(A, \phi_1), I(B, \phi_1)] > D'[I(A, \phi_1), I(A, \phi_2)].$$

Ref. [9] can be viewed as an example for such D', introducing a lighting insensitive comparison function that used gradient features.

In this paper we introduce the idea of leveraging a large reference Library of images that exhaustively samples $I(P_i, \phi)$ over many imaging conditions ϕ for each individual P_i. Data-driven approaches have been employed previously in different problem domains and include [17, 18]. The very recent work in [35] uses a concept similar to the notion of a Library. The main differences are our identity-based Doppelgänger lists and the applicability to a large difference in poses.

Now, say we are given a probe image I_A of person A under unknown imaging conditions and find the most similar image to I_A in the Library. The individual P_i in the Library with the smallest distance to I_A will be a look-alike or *Doppelgänger* of the probe. Furthermore, those images are likely to have been taken under similar imaging conditions because of Eq. 1. For the specific imaging condition of the probe, P_i is the most similar individual in the Library. Looking beyond the Doppelgänger, all individuals in the Library can be sorted based on their distance to the probe, I_A. The sorted Doppelgänger list along with the distances serves as a signature for the probe image. Figure 2 shows an example of two such sorted lists for two probes of the same individual. We argue in this paper that this signature is a good characterization of an individual. This means that if we take a second probe image I'_A of person A, the signature of I'_A will be more similar to the signature of I_A than the signature of an image I_B of another person, even when the images I_A, I'_A are taken under very different imaging conditions and I_B is taken under the same imaging condition as I_A or I'_A.

We explore this idea of using a sorted Doppelgänger list as a signature and evaluate distance functions for comparing the signatures. A subset of the individuals from the Multi-PIE dataset [15] is used as the Library. Our methods are evaluated on a hold-out set of Multi-PIE [15], FacePix [4], and the LFW [12] benchmark.

The following section gives an overview of relevant work. Secs. 3– 5 describe our proposed data driven face-similarity measure and Sec. 6 demonstrates the practicality of our method on Multi-PIE, FacePix, and LFW. Sec. 7 concludes and discusses the prospect of the prosed approach.

2. Related Work

The challenges of large pose variations as well as robustness to illumination and expression changes have been confronted in previous works. In this section we introduce approaches that address these difficulties.

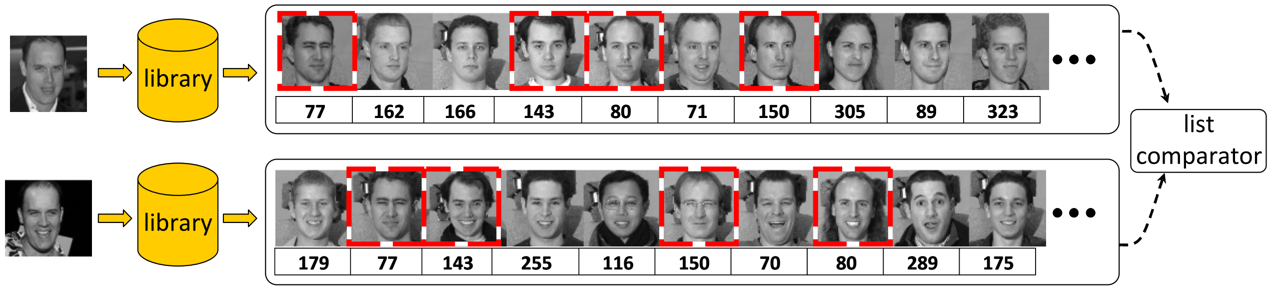

Figure 2. **The pipeline** of our method. Each probe in a query pair is compared to all members of the Library. The comparison results in a ranked list of look-alikes, the first being the most similar to the query. Then, a similarity measure between the two probes is computed by comparing the ranked lists. Here, the probes are from a positive pair in LFW, and the library images are from multi-PIE. The top of the lists have four identities in common, appearing in different poses in each list.

Some methods pursue the idea of constructing the appearance of a face under different imaging configurations in order to facilitate recognition under these different parameters. Georghiades *et al.* [13] showed that illumination changes can form a convex cone. However, changes in pose and expression involve 3D movement and non-rigid changes and can not be modelled by linear transformations in the image space. When 3D data is available, geodesic distances on the face surface can be used to overcome this problem [6]. Attempts were also made to warp 2D data onto 3D models [5]. In addition, some works proposed transforming appearances between poses [2, 24]. Neither of these methods report recognition results, and while reconstruction has worthwhile applications it might not be well suited for biometric scenarios where detailed statistics of the persons appearance matter.

Other methods do not explicitly reconstruct faces but rather focus on a more abstract representation of the underlying training data in order to achieve recognition. Methods like [8, 25] represent the gallery images in terms of manifolds. Ref. [16] leverages the training data by means of a nearest neighbor classifier and employs metric learning in order to compute meaningful distances in the image feature space. Nevertheless, none of the above methods aims at matching faces at *extreme variations of imaging conditions*.

Since its introduction in 2008 Multi-PIE has been used in [7, 28, 33, 34] amongst others. All these approaches restricted the poses to frontal or near frontal. In some, the range of illumination and expression configurations is further restricted to simplify the recognition task. Ref. [28] reports results similar to [33] on face classification on a subset of the five most frontal poses. Cao *et al.* [7] report verification performance on seven frontal poses and a restricted subset of illuminations in addition to neutral expression only.[1]

[1]This restriction was described to us in private correspondence with the authors.

3. A Data Driven Approach for Face Similarity

In this section we detail our method, that was introduced in Sec. 1. Our approach is based on using Doppelgänger lists as face signatures, as summarized in Fig. 2. Given a pair of probes, we compute the look-alike ranked list for each probe from the Library. Then, the similarity between the probes is determined by the similarity between these two lists. This approach stems from the observation that ranked Doppelgänger lists are similar for similar people. This observation is substantiated in Fig. 3.

We calculated the ranked lists for 6000 pairs of probes, having equal number of positive pairs (where the same individual appears in both probe images) and negative pairs. For each pair of lists we counted the location changes of the look-alikes between the lists. For example, if look-alike X was in position 12 in one list and is in position 15 in the second list, this change is counted in cell (12,15). Fig. 3 shows the resulting matrix, divided into bins of 50 by 50 positions. Each row was normalized to represent a probability distribution. Thus, cell (i, j) contains the probability that a look-alike will move from location i to location j. The left plot depicts this probability for positive pairs and the right one for negative pairs.

The peaked diagonal in Fig. 3 (left) supports our assumption that lists computed for a pair of the same identity are similar. In particular, there is a high probability that the look-alikes in the beginning of the list of one probe are at the beginning of the list for the other probe. The peak at the end of the list can be explained by misaligned images that do not match any probes. On the other hand, in Fig. 3 (right), the probabilities are approximately uniform for all possible position swaps. Thus, we conclude that the Doppelgänger lists convey significant information about identity matching.

A major advantage of our method is that direct comparisons between faces are required only in *similar* imaging conditions, where they are actually feasible and informative. Note, that here we use similarity between entire faces of multiple identities to describe people, whereas [20] described an individual according to a collection of similes of

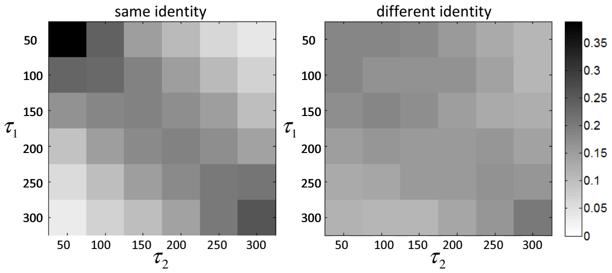

Figure 3. **Probabilities of location changes** between ranked lists of similar and different identities. Positions are binned by 50 positions. These probabilities were computed from 12000 lists, having equal number of matching and non-matching pairs. The diagonal in the left plot supports our assumption that lists computed for a pair of the same identity are similar. In particular, there is a high probability that the look-alikes at the top of the list will be ranked high in both lists. On the contrary, in the right figure the probabilities are more uniform for all possible position swaps.

face parts. Wolf *et al*. [31, 32] utilize a similar idea. However, their ranking based background similarity uses ranked lists of *images*, whereas we rank based on the *identity*. This is a crucial difference and necessary to achieve real pose invariance.

Using Multi-PIE as a Face Library

Currently, the CMU Multi-PIE face database [15] is the most extensive face database in terms of systematic variation of imaging parameters, including expressions, and therefore is the most suitable to act as our Library. Multi-PIE contains images of 337 people recorded in up to four sessions over the span of five months. Over multiple sessions, subjects were imaged under 15 view points (Fig. 4 top), 13 spanning 180° horizontally and two looking from above, imitating a surveillance camera.[2] Strobes were placed at the same locations as the cameras, with an additional three between the two upper cameras, providing 18 illumination conditions. In addition, images were taken under ambient light. Subjects displayed six facial expressions: neutral, smile, surprise, squint, disgust and scream (Fig. 4 bottom). These expressions involve non-rigid movements: eyes open or closed and the mouth varies from closed to widely opened. In total the database contains more than 750,000 images. This multitude of data supports a data driven approach for face matching. Thus, in our experiments images from Multi-PIE formed the face Library.

4. Finding Look-Alikes

In this section we describe how the look-alike ranked lists are retrieved. Each probe is compared to all images in the Library, over all pose, illumination and expression combinations. The Library members are then ordered according to decreasing similarity to the probe. This search could be

[2]In our experiments we use the 13 horizontal view points.

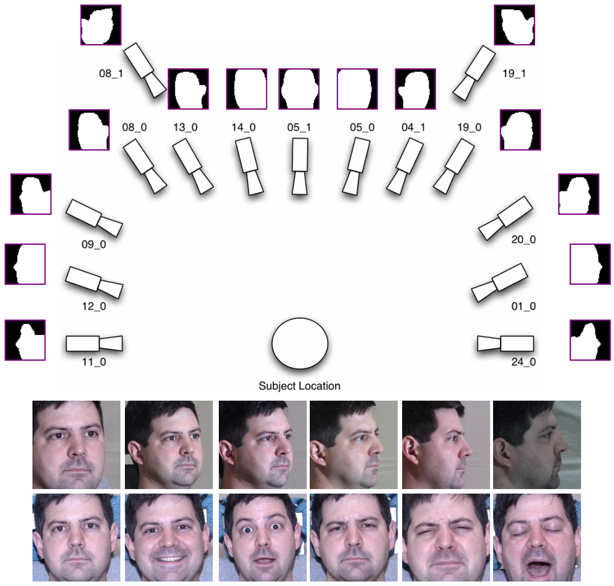

Figure 4. **Multi-PIE**: [Top] The 15 poses available in the Multi-PIE database (figure from [14]). Next to each pose we show the average face support mask for this pose. Masks were obtained by thresholding the illumination variance of the same scene across the 18 strobes, followed by a skin mask. [Bottom] An example of six different poses and the six different expressions in Multi-PIE for the same person.

constrained by a crude estimation of pose or expression.[3] However, we show robust performance even without this step. The face images are aligned depending on the experiment (see Sec. 6).

As discussed in Sec. 1, pixelwise measures are problematic to use across poses. We found that the Structural SIMilarity index (SSIM) [30] yielded the most robust performance across multiple poses. This index was developed to measure similarity between images as a substitute to pixelwise methods such as sum of squared distances. The SSIM measure between pixels in the same location $\mathbf{x}$ in images I_1 and I_2 is given by

$$s_{I_1 I_2}(\mathbf{x}) = \frac{\mu_1(\mathbf{x})\mu_2(\mathbf{x})\sigma_{12}(\mathbf{x})}{[\mu_1^2(\mathbf{x}) + \mu_2^2(\mathbf{x})][\sigma_1^2(\mathbf{x}) + \sigma_2^2(\mathbf{x})]} \quad . \qquad (2)$$

Here, $\mu_i(\mathbf{x}), \sigma_i(\mathbf{x}), i = 1, 2$ are the local mean and standard deviation of a fixed window[4] around $\mathbf{x}$. The value σ_{12} is the correlation of pixels in this window in both images. The values of $s_{I_1 I_2}(\mathbf{x})$ are averaged across all pixel location to produce the final similarity measure. This measure accounts for luminance, contrast and similarity, and ranges between $[0, 1]$, where 1 is achieved for identical images.

Figs. 2 and 5 show examples of look-alike lists computed for several probes. Each list contains an ordering of all N

[3]As an example, [7] estimated crude pose information by comparing to images from Multi-PIE.

[4]In our experiments we used a window size of 9 pixels for image size 50 by 50 pixels.

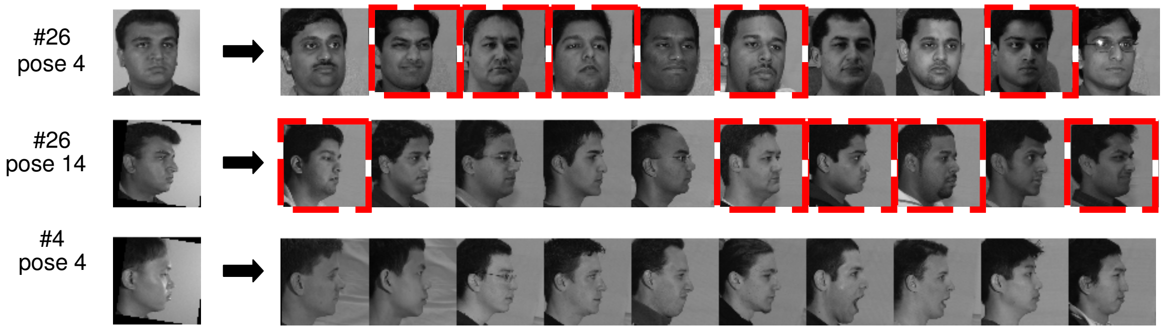

Figure 5. **Doppelgänger lists**: Two probes of the same person (#26) in different poses have five look-alikes that occur in the first ten places. A second person (#4), in the same pose and expression does not have any common look-alikes in the beginning of the list. Red squares mark identities that appear in both lists. Pairwise SSIM between #26 and #4 in the same pose yields higher similarity than the two images of #26 (0.51 vs. 0.25). Probes are from FacePix.

identities in the Library; the figures display the first ten positions. Note that although the comparison was not restricted[5] to specific pose, illumination and expression configurations, the SSIM measure returns look-alikes that are imaged in very similar conditions. This is consistent with the earlier statement of Adini *et al.* [23].

5. Comparing Look-Alike Lists

The output of Sec. 4 is a ranked list of all N distinct identities in the Library for each probe, ordered by decreasing similarity from the probe. The lists, denoted L_1 and L_2 respectively, are permutations of the identity labels in the Library. For simplicity, we assume the labels are numbered from $[1, N]$. In the literature there are a few commonly used statistical methods for comparing permutations such as the Spearman rank, the Spearman Footrule, and the Kendall Tau [21]. The Kendall Tau counts how many pairs of labels switched direction between lists. The Spearman rank and the Spearman footrule are based on counting differences in the positions of the same labels across the two lists. Recently, these methods were extended to accommodate element and position weights [21]. Distance between permutations was also used for finding correspondence in images [3].

We discovered empirically that the similarity value between the probe and the Library instances drops significantly at some point. After this point, the differences between the similarity scores are small and the order between the look-alikes is less informative. This also manifests in the left hand side of Fig. 3, where the look-alikes in the beginning of the list demonstrate a higher tendency to stay in the same position.

Based on these observations, we use the similarity measure of Jarvis and Patrick [19]. For N subjects in the Library, let $\tau_i(n)$ denote the position of subject n in the ordered list L_i. Considering only the first k people in the list where the similarities are significant, the measure is given by (with $[\cdot]_+ = \max(\cdot, 0)$):

$$R(L_1, L_2) = \sum_{j=1}^{N} [k+1-\tau_1(j)]_+ \cdot [k+1-\tau_2(j)]_+ \quad . \quad (3)$$

Fig. 5 shows an example of calculating R for $k = 10$ with probes from the FacePix database. Two probes of the same person (#26) in different poses have five look-alikes that are the same in the first ten positions. A second person (#4), in the same pose does not have any common look-alikes in the beginning of the list. Calculating direct pairwise similarity using SSIM (Eq. 2) yields 0.25 between the two images of #26, and a higher score of 0.51 between the images in the same pose '4'. Thus, in this case the measure R from Eq. (3) has the properties of the desired measure D', as described in Sec. 1.

The methods discussed above ignore all information about the actual similarities $s_{I_1 I_2}$ between the probe and the look-alikes. We view this property as an advantage, because it provides invariance to the scale of the similarities across poses. Nevertheless, future methods might benefit from incorporating the pairwise similarity values into the ranked lists similarity measure.

6. Experimental Evaluation

This section demonstrates the performance of our method on the task of verification. The comparison scores R could be readily used for recognition as well. We show results on Multi-PIE [15], FacePix [4] and LFW [12]. Ten fold cross-validation is performed for all experiments. For Multi-PIE, the dataset was split in ten parts with disjoint identities. The probes (300 positive and negative pairs) were sampled from one set and the Library was built from the remaining nine sets. For FacePix the probes where randomly sampled within the respective pose ranges. For LFW we use the ten folds provided by [12].

[5]Comparing each probe to a large Library may appear to have a significant impact on performance. However, if real-time performance is required, other descriptors may be used, combined with a fast ℓ_2-norm that is sped up with approximate nearest-neighbour techniques [1].

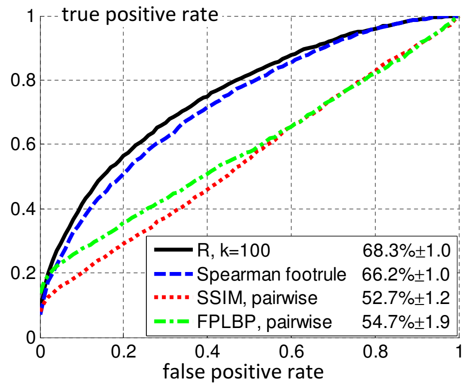

Figure 6. **FacePix: Average ROC curves** obtained by using several similarity measures between the ranked lists across all poses. The measure in Eq. (3) is used with $k = 100$.

We emphasize that to the best of our knowledge face verification that uses the *entire* range of poses, illuminations, and expressions has not been reported neither on Multi-PIE, *e.g.* [7, 28], nor on FacePix, *e.g.* [10]. Previous results are limited to subsets of the conditions and exclude some of the hardest conditions (*e.g.* profile views [7]).

We compare our Doppelgänger list based face-similarity using $R(L_1, L_2)$ with one of the state of the art descriptors used for faces (FPLBP [32]) as well as SSIM for direct pairwise comparison of images. In addition we compare our results on LFW to the performance reported by [16, 32] for the low-level features.

6.1. Rank comparison measures

We tested different measures for comparing ranked lists [21]. Fig. 6 shows average ROC curves obtained by using several similarity measures between the ranked lists across the full range of poses in FacePix. Results are reported for verification accuracy at equal error rate (EER). Decreased performance for Spearman footrule, compared to R in Eq. (3) indicates, that the top of the list (*e.g.* top 100) is more important than identities further down the list. Spearman-rank and Kendall-tau perform comparable to the Spearman footrule. We use R in subsequent experiments.

6.2. Verification Across Large Variations of Pose

In the first experiment we use ten test sets selected from FacePix, each consists of 500 positive and 500 negative verification pairs. Poses were restricted according to the experiment, and can range from $-90°$ to $90°$. We use the face and fiducial-point detector from [11] to align the FacePix probes to the Multi-PIE based Library. Note that this alignment was only performed for images in the *same or similar* poses, thus circumventing the flawed procedure of trying to align images across a large pose range. Each probe was similarity-aligned to the canonical coordinates of the corresponding Multi-PIE pose, which was assumed to be

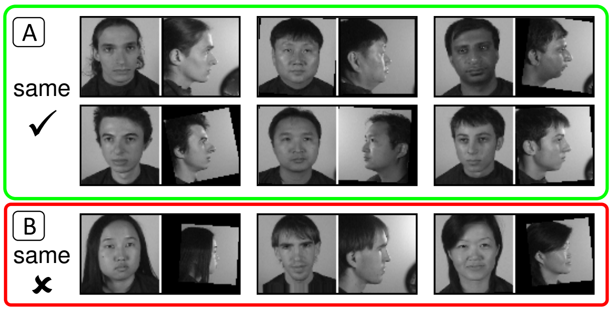

Figure 7. **FacePix example pairs** that were classified correctly and incorrectly using Eq. (3). Shown are pairs of correctly (A) and incorrectly (B) classified pairs of the same identity.

probe 1	probe 2	Pairwise FPLBP [32]	Look-alike measure
$-30°$ to $30°$	$-30°$ to $30°$	$72.0\% \pm 2.0$	$74.5\% \pm 2.6$
$-90°$ to $90°$	$-90°$ to $90°$	$54.7\% \pm 1.9$	$68.3\% \pm 1.0$
$-10°$ to $10°$	angle$>|70|$	$51.1\% \pm 1.7$	$66.9\% \pm 1.0$

Table 1. **FacePix across pose**: Summary of classification accuracy at EER with test sets of various pose ranges. Our method is compared to FPLBP as used in [32] on image pairs. As expected, the performance of the pairwise measure decreases with increased pose variations and approaches chance, whereas the performance of our measure (R, $k = 100$) remains relatively stable.

known. However, our experiments showed that the matching of probes to the Library could also function as a pose estimator and obviate this assumption.

Robustness to Pose Variation. From Tab. 1 it is apparent that our method's performance remains high across the full range of poses, whereas the comparison metric FPLBP[6] (one of the best performing single measures on pairs of faces [16, 32]) drops significantly. The same decrease was reported on other methods [27]. Adding profile poses to the experiments slightly degrades the performance of our method, but it still copes with the increased difficulty. The proposed method does not rely on alignment across poses, which makes it more robust in cases where it is impossible to align faces relative to each other (*e.g.* frontal to profile).

Fig. 7 depicts a few examples of correctly and incorrectly classified pairs. It is noticable that especially for profile poses the fiducial point detector [11] does not perform well which leads to misalignment and therefore noisy Doppelgänger lists. Better alignment can certainly help to improve performance in those cases. Note that our method *only* relies on good registration of a pair of images within a small range of poses (in the Multi-PIE based library $\pm 15°$).

Across all poses on Multi-PIE. Again, our face-similarity proves itself robust across the full range of poses, illumination, and expression conditions. Faces were detected us-

[6] In our experiments FPLBP performed better than TPLBP, and was used with the χ^2-distance.

Method	TPLBP	FPLBP	SIFT	look-alike
Accuracy at EER	69.2%	68.2%	69.1%	**70.8%**

Table 2. **LFW**: Comparison of the proposed look-alike face-similarity with other pairwise feature measures as reported in [32].

ing [11] and bad detections were rejected. The classification performance using ten fold cross-validation is $76.6\% \pm 2.0$ (both FPLBP and SSIM on direct image comparison perform near chance). To the best of our knowledge these are the first results reported across *all* pose, illumination and expression conditions on Multi-PIE.

6.3. Results on LFW

In our final experiment we show the ability to generalize to unseen real world images. We report improved performance of our face-similarity on LFW [12] in comparison with results reported for low-level feature descriptors that aim to solve the same problem (*e.g.* LBP, Gabor, TPLBP, FPLBP, SIFT). We use the LFW-a aligned faces [32] and aligned the Multi-PIE Library faces in a similar way using a commercial fiducial point detector.

As before we use Eq. (3) with $k = 100$ and SSIM as the measure to compute the look-alike lists. The performance on LFW using our face-similarity is $70.8\% \pm 1.5$ (see Tab. 2). This outperforms other metrics that calculate a similarity score for faces. Wolf *et al.* [32] report performance around 69% using TPLBP or SIFT (LBP, Gabor, FPLBP are reported with slightly lower scores), and in our experiments the SSIM measure performed at $69\% \pm 2$. This signifies an improvement on LFW, while keeping in mind that the main advantage in our face-similarity measure is the robustness in comparisons across poses, which is not captured by the LFW benchmark.

Fig. 8 shows examples from this experiment, for both correctly and incorrectly classified pairs. The top box shows pairs of the same person that were classified correctly. Note the high variability of pose, expression, illumination among the pairs. Some positive pairs that were classified incorrectly exhibit extreme variations in imaging conditions, as well as significant occlusions.

This experiment underlines the generalization of our method. As shown in experiments on FacePix and Multi-PIE the real strength lies in generalization across large pose ranges, where direct pairwise comparisons would fail. Similar to [32] our method could be combined with other state of the art methods to combine the advantages of standard descriptors and our very different approach.

7. Discussion

To the best of our knowledge, we have shown the first verification results for face-similarity measures under truly unconstrained expression, illumination and pose, including full profile, on both Multi-PIE and FacePix.

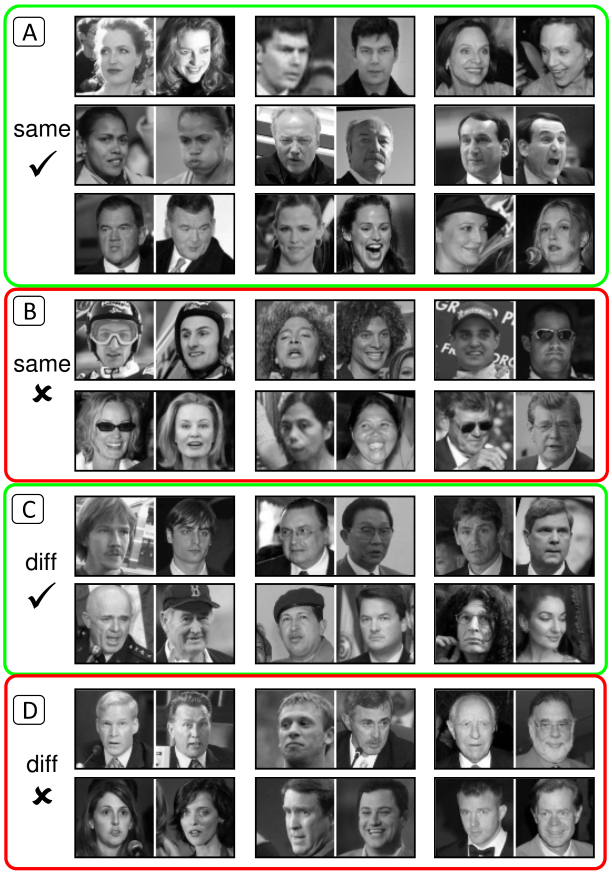

Figure 8. **Example pairs** that were classified correctly and incorrectly in LFW using Eq. (3). Shown are pairs of correctly (A) and incorrectly (B) classified pairs of the same identity, and correctly (C) and incorrectly (D) classified pairs of different identities.

The advantages of the ranked Doppelgänger lists become apparent when the two probe images depict faces in very different poses. In this scenario many descriptors [16, 32] fail, as shown in Tab. 1. Even fiducial point based descriptors such as SIFT or affine descriptors are expected to struggle across such extreme ranges of poses. Methods like the attributes based classifier [20] may potentially cope with large pose variations, provided that the attribute classifier is trained specifically for separate pose ranges. Our method does not require explicit training and is able to cope with large pose ranges.

It is straightforward to generalize our method to an even larger variety of imaging conditions, by adding further examples to the Library. No change in our algorithm is required, as its only assumption is that imaging conditions similar to the ones used for the probe exist in the Library. The robustness to this assumption has been proven in the FacePix and LFW experiments, where the exact pose is not contained in the Library.

We expect that a great deal of improvement can be achieved by using this powerful comparison method as an additional feature in a complete verification or recognition pipeline where it can add the robustness that is required for face recognition across large pose ranges. Furthermore, we are currently exploring the use of ranked lists of identities in other classification domains.

Acknowledgments

We thank Eric Christiansen for his discussions. This work was supported in part by ONR MURI Grant #N00014-08-1-0638 and UCSD FWGrid Project, NSF Grant #EIA-0303622. Tali Treibitz is an Awardee of the Weizmann Institute of Science – National Postdoctoral Award Program for Advancing Women in Science.

References

[1] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Wu. An optimal algorithm for approximate nearest neighbor searching. *Journal of the ACM*, 45(6):891–923, 1998.

[2] S. Avidan and A. Shashua. Novel view synthesis in tensor space. In *Proc. IEEE CVPR*, 1997.

[3] D. Bhat and S. Nayar. Ordinal measures for image correspondence. *Trans. IEEE PAMI*, 20(4):415–423, 2002.

[4] J. Black, M. Gargesha, K. Kahol, P. Kuchi, and S. Panchanathan. A framework for performance evaluation of face recognition algorithms. In *ITCOM*, 2002.

[5] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *Trans. IEEE PAMI*, pages 1063–1074, 2003.

[6] A. Bronstein, M. Bronstein, and R. Kimmel. Expression-invariant 3D face recognition. In *Audio-and Video-Based Biometric Person Authentication*, 2003.

[7] Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning-based descriptor. In *Proc. IEEE CVPR*, 2010.

[8] H. Cevikalp and B. Triggs. Face Recognition Based on Image Sets. In *Proc. IEEE CVPR*, 2010.

[9] H. Chen, P. Belhumeur, and D. Jacobs. In search of illumination invariants. In *Proc. IEEE CVPR*, 2000.

[10] K. W. Cheung, J. Chen, and Y. S. Moon. Pose-tolerant Non-frontal Face Recognition using EBGM. In *Proc. Intl. Conf. Biometrics: Theory, Applications and Systems*, 2008.

[11] M. Everingham, J. Sivic, and A. Zisserman. “Hello! My name is... Buffy” – automatic naming of characters in TV video. In *Proc. BMVC*, 2006.

[12] B. Gary, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[13] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *Trans. IEEE PAMI*, 23(6):643–660, 2001.

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

[15] R. Gross, I. Matthews, J. F. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *Proc. IEEE Conf. on Automatic Face and Gesture Recognition*, 2008.

[16] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? Metric learning approaches for face identification. In *Proc. ICCV*, 2010.

[17] A. Halevy, P. Norvig, and F. Pereira. The unreasonable effectiveness of data. *Intelligent Systems, IEEE*, 24(2):8–12, 2009.

[18] J. Hays and A. Efros. IM2GPS: estimating geographic information from a single image. In *Proc. IEEE CVPR*, 2008.

[19] R. Jarvis and E. Patrick. Clustering using a similarity measure based on shared near neighbors. *Trans. IEEE on Comp.*, pages 1025–1034, 1973.

[20] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. Attribute and Simile Classifiers for Face Verification. In *Proc. ICCV*, 2009.

[21] R. Kumar and S. Vassilvitskii. Generalized distances between rankings. In *Proc. Int. Conf. on world wide web*, 2010.

[22] A. Martinez and R. Benavente. The AR face database. Technical Report 24, CVC Technical Report, June 1998.

[23] Y. Moses, Y. Adini, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction,. In *Proc. ECCV*, 1994.

[24] J. Ni and H. Schneiderman. Face view synthesis across large angles. In *Proc. Workshop on Faces and Gestures, ICCV*, 2005.

[25] S. Park and M. Savvides. An extension of multifactor analysis for face recognition based on submanifold learning. In *Proc. IEEE CVPR*, 2010.

[26] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *Trans. IEEE PAMI*, 22:1090–1104, 2000.

[27] N. Pinto, J. DiCarlo, and D. Cox. How far can you get with a modern face recognition test set using only simple features? In *Proc. IEEE CVPR*, 2009.

[28] D. Ramanan and S. Baker. Local distance functions: A taxonomy, new algorithms, and an evaluation. In *Proc. ICCV*, 2010.

[29] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination and expression (PIE) database. In *Proc. IEEE Conf. on Automatic Face and Gesture Recognition*, 2002.

[30] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *Trans. IEEE Image Processing*, 13(4):600–612, 2004.

[31] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In *Proc. IEEE CVPR*, 2011.

[32] L. Wolf, T. Hassner, and Y. Taigman. Similarity scores based on background samples. In *Proc. ACCV*, 2009.

[33] J. Wright and G. Hua. Implicit elastic matching with random projections for pose-variant face recognition. In *Proc. IEEE CVPR*, 2009.

[34] A. Yang, A. Ganesh, Z. Zhou, S. Sastry, and Y. Ma. Fast l1-minimization algorithms and an application in robust face recognition: a review. In *Proc. ICIP*, 2010.

[35] Q. Yin, X. Tang, and J. Sun. An associate-predict model for face recognition. In *Proc. IEEE CVPR*, 2011.