RPM: Random Points Matching for Pair-wise Face-Similarity

M. Saquib Sarfraz*¹

saquib.sarfraz@kit.edu

Muhammad Adnan Siddique*²

siddique@ifu.baug.ethz.ch

Rainer Stiefelhagen¹

rainer.stiefelhagen@kit.edu

¹ Institute for Anthropomatics

Karlsruhe Institute of Technology

Karlsruhe, Germany

² Swiss Federal Institute of Technology

(ETH) Zurich

Zurich, Switzerland

Abstract

Matching face image pairs based on global features or local analysis on some points found using a key point or fiducial point detector becomes prohibitively difficult in realistic images when there are large pose, lighting, expressions and imaging differences. We develop a new approach that automatically and reliably finds well-matched and useful corresponding points, referred to as homologous points, from randomly initialized points on the two probe images under unrestricted image variations. The procedure obviates the need of using key or fiducial point detector and the over restrictive requirement of image alignment. We then propose a new pair-wise similarity metric that combines the strength of the useful parameters found during the random point matching and the similarity computed using a local descriptor around the homologous points. Our results in a face verification setting on two challenging datasets ('Labelled Faces in the Wild' and FacePix) under large pose, expression and imaging variations, show improved performance over the state-of-the-art methods for pair-wise similarity.

1 Introduction

Pair-wise face verification strives to determine whether two probe images belong to the same person. With the increasing demand of face recognition/verification in real-world applications, such as access control, face tagging on the social websites [1], person retrieval in videos [2] etc., the challenges confronted by the verification algorithms include not only dealing with large pose and lighting variations, but also very different imaging conditions e.g. different imaging device or resolution, sensor noise, varied expressions, lens distortion and occlusions.

Face verification algorithms generally require image representation followed by matching. For the representation, the current methods typically rely on aligning the two images such that the effect of pose and other geometric deformations are minimized. A feature analysis is then performed for matching, either globally on the whole face, or locally on some points typically found by a fiducial point detector. The over restrictive requirement of image alignment, especially for global feature analysis, is not trivial to meet. Besides the need to

Figure 1: Point features matching in a pair of positive images: (a) with SIFT detector and descriptor; (b) with RPM (our approach).

know the pose a priori, it requires 2D warps or 3D model-fitting which in turn depends on reliably detected fiducial points. Therefore, the performance of a verification algorithm is inevitably limited by the possibility of detecting reliable fiducial points.

In general, even the state-of-the-art fiducial or key point detectors do not guarantee to find all the required fiducial points. SIFT and similar detectors may fail to provide useful points for matching, especially when the probe images exhibit wide imaging differences. The points found do not guarantee to correspond to the same physical location in the two images and thus result in poor matching correspondence. Figure 1 a. depicts this problem by showing the matches found by employing the SIFT key point detector and matching engine [14]. The image pair shown is comprised of a high resolution frontal gallery face and the probe face (detected from a poor resolution surveillance video of the same individual). Due to the underlying imaging differences, the SIFT detector has failed to provide good corresponding points belonging to the same physical location on the two images. Fiducial point detectors pose similar behaviour, particularly in case of wide pose differences (e.g. when matching a frontal and profile view).

In this paper we shift from this over reliance of key point detectors and propose a novel method to automatically obtain well-matched points across the two images, subsequently referred to as the homologous points (HPs). It is based on randomly initializing enough points on the detected face windows and then follow up with a powerful mutual information assisted matching strategy. The approach lifts the restrictions, noted in the preceding paragraph, of the current key/fiducial point detectors as it is more probable to yield well matched pairs of points (on the same physical location on the face) from the randomly initialized points by using a meaningful iterative matching strategy. The procedure, as explained in the next sections, is such that it automatically finds pairs of points that not only correspond well but are also meaningful in establishing the similarity of the face in the image pair. To finally match the image pair we propose a new similarity metric that combines the useful parameters found during random point matching and the similarity scores computed by using a local descriptor on these homologous points. In section 2 we survey some related work, section 3 details our approach of random point matching, section 4 proposes the new similarity metrics and section 6 concludes the paper by providing extensive experimental evaluations of the proposed method in section 5.

2 Related Work

In this section we review the literature related to the computation of pair-wise similarity for the task of face verification. The current approaches can be largely categorized in two direc-

tions; those that compute direct similarity based on proposing new representation methods [0, 1, 15, 18], and those that employ a learning based framework to learn the similarity functions [2, 3, 4, 14, 19, 20]. Both rely on a point detector to either compute direct similarity by using some feature description method or learn the similarity scores or some parts on the face to later use for recognition. While our method falls in the first category, it is equally useful for the learning based approaches.

Descriptor based methods have been proven effective for the face representation. Ahonen et al. [10] proposed Local Binary Pattern (LBP) to describe the micro-level patterns on the face. Many LBP variants since then have been proposed. Wolf et al. [18] proposed a Three Point LBP (TPLBP) and Four point LBP (FPLBP) and showed the best performance of these descriptors in a pair-wise face similarity verification task. Similarly other descriptors e.g. Bioinspired Gabor [2] and SIFT (on fiducial points) have also been employed and the performance has been reported in [32]. A more recent interesting work is of Schroff et al. [19] that employs a data driven approach for computing pair wise similarity. They use the comparisons of each image in the pair with a large set of library images and use the similarity of these sorted ranked lists as the metric. The approach performs very good in the presence of large pose and illumination variations, however it is slow as it requires many thousands of matches for each pair while intrinsically requiring the good alignment between the probes and the library images. The methods based on directly matching the extracted features have their limitations due to the large appearance variations.

Learning based approaches try to model these variations while computing similarity scores. Cao et al. [8] learned the different descriptors (e.g. LBP, SIFT) by using unsupervised learning techniques to come up with a discriminative face representation. Sarfraz et al. [14] learned the similarities computed using GLOH features in a probabilistic framework to explicitly model the variations due to pose differences. Some recent methods e.g. [9] employ a metric learning approach that learns an objective function using a discriminant classifier on a set of positive and negative examples. More recent approaches such as Wolf et al. [18], Yin et al. [20] and Berg and Belhumer [4] learn specialized parts classifiers on the pair taking advantage of a reference set of images. An interesting shift from these is the approach used by Kumar et al. [11] that learns the attributes called similes (such as hair color, age, gender etc.) from the images. The approach is attractive as it is not dependant on the underlying imaging conditions e.g. pose, illuminations etc. but the problem of learning these attribute classifiers still poses many challenges from the unrestricted images.

All of these methods rely on the properly registered/aligned images. For a large degree of pose variation, alignment involves 2-D warps or a 3-D [1, 5] model fitting based on the detected key-points on the face.

3 Point Matching Formulation

As detailed in the preceding section almost all of the current methods rely on a key/fiducial point detector either for alignment or matching the points directly. Here, we argue that this key/fiducial point detector may not guarantee a good correspondence in the presence of large pose and other imaging variations especially when the underling image modalities are different due to , for example, different image sensors. This is largely because these detectors try to find the specific regions on the face and thus become image specific. We on the other hand propose to find the corresponding point pair (homologous points) that falls on the same physical location from a dense set of randomly initiated points. As detailed in the following

Figure 2: Random Points Initialization and Matching between a pair of face images, followed by outlier rejection: (a) A set of randomly initialized points; (b) Random points matched with normalized mutual information, ambiguities exist; (c) Ambiguities have been removed and the points are uniquely matched; (d) Outliers have been rejected.

sections, these pairs can directly provide useful image similarity information. While we use these points to establish direct pair-wise similarity, nonetheless, all of the current methods may benefit by using these newly established correspondences. The following sub-sections sequentially present the various steps in our methodology.

3.1 Random Points Initialization and Matching

A set of uniformly distributed random points is initialized in both the images, as shown in Figure 2a. If the images are well-aligned frontal poses of the same person, the points would tend to fall in more or less the same local regions of the face. However, variations in the facial expressions would still not guarantee it. In case of wide variations in poses and expressions, this possibility is obviously unexpected. To cater for variations in poses, facial expressions, and other intrinsic differences in the imaging sensors (such as camera tilts), the RPM methodology inherently allows the potential matching of a random point in one image with any random point in the other image.

We use normalized mutual information (nMI) as a similarity measure in the matching process. It is computed on the candidate regions ‘C’ of predefined size centred on each random point in both the face images. The normalized mutual information of the i^{th} candidate regions, C_A^i and C_B^i , in the two face images, A and B respectively, is computed as follows:

$$\text{nMI} \left(C_A^i, C_B^j \right) = \frac{H(C_A^i) + H(C_B^j)}{H \left(C_A^i, C_B^j \right)} \quad (1)$$

$i, j \in \{1, 2, \dots, N\}$, while N is the total number of randomly initialized points. $H(C_A^i, C_B^j)$ is the joint Shannon Entropy of the two candidate regions, while $H(C_A^i)$ and $H(C_B^j)$ are the marginal entropies. These entropies can be computed using the joint and marginal histograms of the candidate regions. The random points across the two images are then matched

on the criterion that the higher the mutual information between two candidates, the more likely they are to represent the same local information on the face. Hence, all the points in one image are matched to some other point in the other image. The following equation summarises this matching step:

$$m^i = \arg \max_j \text{nMI} \left(C_A^i, C_B^j \right) \quad (2)$$

where m^i is the specific point in face image B that is matched to the point i in face image A, subsequently denoted with $i \rightarrow m^i$. The search space for j during the maximisation process in equation 2 can be constrained to a local region, i.e. only those points that fall within a predefined local region are analysed instead of all the points throughout the image, and the match is established with the one that offers the highest mutual information.

Figure 2b shows the matched points. It can be seen in the figure that occasionally, multiple points in one image get matched to the same point in the other image. In other words, for each m^i , there may be multiple instances of point i , i.e. $\{i_1, i_2, \dots\} \rightarrow m^i$. These ambiguous matches, though undesired, are nonetheless not unexpected. A statistical measure of similarity like mutual information cannot in itself guarantee a ‘one-to-one’ correspondence among the candidate regions. To alleviate these ambiguous matches, we adopt a sequential approach of retaining only the particular pair that exhibits the maximum MI in comparison to other matched pairs. Mathematically, from among the set of matched points, $\{(i_1 \rightarrow m^i), (i_2 \rightarrow m^i), \dots\}$, the one (i_r) that offers the highest MI is retained as a pair

$$i_r = \arg \max_{i_k} \text{nMI} \left(C_A^{i_k}, C_B^{m^i} \right) \quad (3)$$

$$i_r \rightarrow m^{i_r}$$

and the remaining points are matched to those points in the other image that offer the second highest mutual information. However, this may not remove all the ambiguous matches as yet. The process in equation 3 is repeated until all pairs are uniquely matched, as shown in Fig 2c.

3.2 Outlier Rejection with Recursive Model Fitting

The matching strategy in the preceding subsection does not ensure that the matched points would belong to the same physical point in the two faces, as the points themselves have originated randomly. At the same time, we can appreciate that most of the matched points tend to belong to more or less the same local neighbourhood. This section delineates a simple but effective method to eliminate the outliers in our case.

We assume that there exists a geometric model that would map each random point to its match. For a single point and its match, the model is essentially a straight line segment joining the two, characterised by the two parameters of gradient and length. Let (x_A^i, y_A^i) be the coordinates of the i th point in image A, and (x_B^i, y_B^i) to be the coordinates of its match in image B. We compute the gradient of the line segment joining the two points, i.e. g_i . Next we group all the pairs, whose gradients differ from g_i within a certain pre-defined tolerance τ , into a set $\mathbf{G}_i$, as follows:

$$\mathbf{G}_i = \{k : |g_k - g_i| \leq \tau\} \quad (4)$$

Figure 3: Homologous points obtained in a negative pair of face images.

Recursively, we obtain $\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_N$. From among these sets, the set with the highest cardinality is retained. The pairs in this particular set do not contain outliers and represent well-matched local face regions, as shown in Fig 2d. It needs to be appreciated here that we can build more complex geometric models involving multiple pairs of matched points, however, a simple straight line gradient-based outlier rejection suffices for our case.

3.3 Area-based Optimisation to obtain Homologous Points (HPs)

Since the points were randomly initialized, as yet the points in a matched pair only establish correspondence between similar face regions among the two images. The localization of the points needs to be further improved so that the matched points represent physically identical locations. We use an area-based implementation of normalized MI to refine the physical positions of these points.

$$\Delta T_{xy}^{i_r} = \arg \max_{(\Delta T_{xy})} n \text{MI} \left(C_A^{i_r}, C_B^{m^{i_r} + \Delta T_{xy}} \right) \quad (5)$$

For each point i_r in image A, the physical location of its matched point m^{i_r} in image B is translated by $\Delta T_{xy}^{i_r}$. This translation allows a further improvement towards the objective of having matched points that originate at identical physical face locations.

The RPM methodology thus leads us from completely random points to homologous points, as shown in Fig 1b. The correspondences provided by RPM may not only be used for direct image matching but also for many other applications such as pose estimation, alignment etc.. We have, for example, used some of the initial ideas presented here in our prior work [14] on the registration of multi-modal satellite imagery.

To signify the strength of the presented methodology, Fig 3 shows the homologous matches in case of a negative pair of images. It can be seen here that despite the inherent differences, RPM has been successful in locating homologous points even with the random initializations. RPM thus revokes the customary need to invoke a point feature detector.

The next section extends the use of RPM towards formulating appropriate similarity metrics to classify an image pair as either positive or negative.

4 Similarity Metrics for Pair-wise Similarity

This section proposes direct metrics of similarity to classify an image pair as positive or negative. These measures are developed on the parameters lent by the RPM methodology in section 3.

The *number of HPs* obtained for a pair relative to the number of random points initialized, n , behaves as a potential metric. Generally, for positive pairs, n is higher than for negative

Figure 4: Average ROC curves for different similarity metrics obtained on 10 test runs on a single fold of the LFW dataset. The accuracies mentioned are at the equal-error-rate (EER).

pairs. This argument has been empirically validated. Fig 4 shows the average ROC curve for 10 runs of the RPM for this metric on a single fold of the ‘Local Faces in the Wild (LFW) [10]’ dataset.

The *variance of the model parameters*, σ , is another potential metric. In case of a negative pair of face images, the gradients as computed in section 3.2, generally exhibit more variance than for a positive pair. The local facial regions in a negative pair are naturally more different than for a positive pair; therefore, the locally computed model parameters would likely show more variance.

The HPs provided by the RPM formulation provide a useful correspondence at a local level among a pair of images. We employ a local descriptor ‘*Local Energy Shape Based Histogram (LESH)*’ [13] because of its superior performance as compared to other commonly used local facial descriptions. For each matched pair, we compute the cosine similarity of the respective LESH vectors. Subsequently, the average of these local similarities ‘ L ’, as given in equation 6, defines our third similarity metric for the image pair.

$$L = \frac{1}{n} \sum_{i=1}^n S(\mathbf{L}_A^i, \mathbf{L}_B^i) = \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{L}_A^i \cdot \mathbf{L}_B^i}{|\mathbf{L}_A^i| |\mathbf{L}_B^i|} \quad (6)$$

where $\mathbf{L}_A^i$ and $\mathbf{L}_B^i$ are the i^{th} LESH vectors in image A and B and $S(\cdot)$ is the cosine similarity between the two. This serves as a very useful similarity metric, as evident from the corresponding ROC in Fig 4. L varies between 0 and 1, with a higher value representing more similarity.

4.1 Q-score

As a final similarity metric for the image pair we propose the following quantitative similarity score (hereafter referred to as the Q -score) that attempts to complement the geometric model information with the local image description from LESH.

$$Q = \exp\left(\frac{(L+1) \times n}{\sigma + n}\right) \quad (7)$$

Higher values of Q -score represent a higher pair-wise face-similarity. In case of an identical pair of images, we expect L to be close to 1, σ to approach 0 and n to approach 1; thus,

Figure 5: Average ROC curves for RPM on different pose groups of FacePix. The accuracies mentioned are at the equal-error-rate (EER).

$Q \rightarrow e^2$. In case of starkly dissimilar images, $Q \rightarrow 1$ as both n and L approach 0 while σ takes high values. The ROC for the Q -score is also shown in Fig 4. It not only performs better than each of the metrics discussed earlier but is also quite stable with respect to the random initializations.

Since RPM invokes random points, the repeatability of the results has to be verified for the same set of pairs. Figure 4 provides average results of 10 runs of the RPM, on one of the folds of LFW, and the standard deviation at the equal-error-rate (EER). The standard deviations are small for each metric, confirming their usefulness. Q -score gave the least deviation that further supports our argument towards its use as a pair-wise face similarity metric.

5 Results

In this section we provide results of our method on two challenging datasets, FacePix [19] and LFW [10]. We report a quantitative comparison with recent state-of-the-art methods that use a pair-wise face similarity for verification.

For the experiments we do not perform any image alignment, the only care has been taken in initializing points so that they fall approximately on the face region as expected from a typical face detector output. We randomly initialize 100 points on each image and the size of candidate region ‘ C ’ around each point is taken to be 32×32 . The tolerance τ in recursive model fitting is set to be 0.03.

FacePix is used to test the performance under wide variations in the poses (180 images of each person in continuous pose variations from -90° to $+90^\circ$). For quantitative comparisons we follow the protocol in [19]. We prepared three groups of test data with respect to pose variations, as shown in Table 1. In each group, we sample randomly 10 test sets each comprising of 500 positive probe pairs and 500 negative probe pairs. These probe pairs were randomly selected, to allow for identity diversity besides pose variation. We report average results in Table 1 for each group. We compare our method with one of the best performing direct pair-wise similarity metric FPLBP [19] and a recent pose invariant data driven approach [15] that reports the best results so far on a direct pair-wise similarity verification task on these datasets.

Our approach provides very good results, even in case of very wide pose variations, while

Probe Image 1	Probe Image 2	FPLBP[19]	Look-alike[15]	RPM: Q -score
-10° to $+10^\circ$	-10° to $+10^\circ$	-	-	$95.6\% \pm 0.8$
-30° to $+30^\circ$	-30° to $+30^\circ$	$72.0\% \pm 2.0$	$74.5\% \pm 2.6$	$75.7\% \pm 1.5$
-10° to $+10^\circ$	angle $> 70^\circ $	$51.1\% \pm 1.7$	$66.9\% \pm 1.0$	$67.2\% \pm 1.7$

Table 1: Performance on FacePix: Comparison of classification accuracies at EER over different pose groups

Method	TPLBP	FPLBP	SIFT	Look-alike	RPM: Q -score
Classification Accuracy:	69.2%	68.2%	69.1%	70.8%	71.6%

Table 2: Performance on LFW: Comparison of classification accuracies at EER, with state-of-the-art pair-wise similarity metrics as reported in [15, 19]

for relatively less pose variations (as expected in a typical verification setting) our method performs exceptionally well. The corresponding average ROC curves are shown in Fig 5.

We also report results on the LFW dataset, on the ten folds as provided in [19]. We achieve $71.6\% \pm 2.3$ average classification accuracy at EER on the 10 sets, which is an improvement in comparison with the other methods that computes a direct pair-wise face similarity, as shown in Table 2.

As can be seen, our method performs better than the approaches that are based on direct image comparison on low-level feature descriptions e.g. LBP, Gabor (C1), SIFT (on fiducial), FPLBP and TPLBP, while it performs comparable (slightly better) than a recent data driven approach [15]. However, it should be appreciated that while we base our scores on direct comparison of the image pair, the approach in [15] uses a strictly aligned large library comparison in order to compute the similarity between two images. We, therefore, still offer better performance across large pose differences while having a large computational advantage.

6 Discussion and Conclusion

The RPM formulation, to the best of our knowledge, is the first attempt at generating reliable homologous points among a pair of face images from completely random points, obviating the need to invoke a key/fiducial point detector as traditionally required before performing a local or global pair-wise feature analysis. It lends us useful metrics for pair-wise face similarity verification, such as the proposed Q -score which performs better than some of the state-of-the-art methods. Moreover, our work helps in setting the new perspective that the non-trivial requirement of image alignment (for an appropriate representation prior to matching) can be circumvented.

We aim to extend this work in two directions. In this paper, the different metrics originating from the RPM formulation have been used to build a direct measure of similarity. We intend to investigate the potential use of these metrics in training a classifier as well. Secondly, as the RPM formulation provides well-matched homologous points even in case of wide variations in imaging conditions, it can be used for pose-estimation and image alignment itself. We expect that the RPM formulation may provide a general solution to various machine vision problems.

Acknowledgements

This work was partially funded by the German Federal Ministry of Education and Research (BMBF) under contract no. 01ISO9052E (PaGeVi) and 13N12063 (MisPel). The views expressed herein are the authors' responsibility and do not necessarily reflect those of BMBF.

References

- [1] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face recognition with local binary patterns. In *ECCV*, pages 469–481. Springer, 2004.
- [2] Akshay Asthana, Michael Jones, Tim Marks, Kinh Tieu, and Roland Goecke. Pose normalization via learned 2d warping for fully automatic face recognition. In *BMVC*, pages 127–1, 2011.
- [3] V Balasubramanian, J Ye, and S Panchanathan. Biased manifold embedding: A framework for person-independent head pose estimation. In *CVPR*. IEEE, 2007.
- [4] Thomas Berg and Peter N Belhumeur. Tom-vs-pete classifiers and identity-preserving alignment for face verification. In *BMVC*, volume 1, page 5, 2012.
- [5] Volker Blanz and Thomas Vetter. Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1063–1074, 2003.
- [6] Zhimin Cao, Qi Yin, Xiaoou Tang, and Jian Sun. Face recognition with learning-based descriptor. In *CVPR*, pages 2707–2714. IEEE, 2010.
- [7] David Cox and Nicolas Pinto. Beyond simple features: A large-scale feature search approach to unconstrained face recognition. In *Automatic Face & Gesture Recognition (FG)*, pages 8–15. IEEE, 2011.
- [8] Mark Everingham, Josef Sivic, and Andrew Zisserman. Hello! my name is... buffy—automatic naming of characters in tv video. 2006.
- [9] Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid. Is that you? metric learning approaches for face identification. In *ICCV*, pages 498–505. IEEE, 2010.
- [10] Gary B Huang, Marwan Mattar, Tamara Berg, Eric Learned-Miller, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008.
- [11] Neeraj Kumar, Alexander Berg, Peter N Belhumeur, and Shree Nayar. Describable visual attributes for face verification and image search. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(10):1962–1977, 2011.
- [12] Jakob Verbeek Cordelia Schmid Matthieu Guillaumin, Thomas Mensink. Automatic face naming with caption-based supervision. In *CVPR*, pages 1–8, 2008.
- [13] M Saquib Sarfraz and Olaf Hellwich. Head pose estimation in face recognition across pose scenarios. *VISAPP (1)*, pages 235–242, 2008.

- [14] M Saquib Sarfraz and Olaf Hellwich. probabilistic learning for fully automatic face recognition across pose. *Image and Vision Computing*, 28(5):744–753, 2010.
- [15] Florian Schroff, Tali Treibitz, David Kriegman, and Serge Belongie. Pose, illumination and expression invariant pairwise face-similarity measure via doppelgänger list comparison. In *ICCV*, pages 2494–2501. IEEE, 2011.
- [16] Muhammad A Siddique, M Saquib Sarfraz, David Bornemann, and Olaf Hellwich. Automatic registration of sar and optical images based on mutual information assisted monte carlo. In *Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1813–1816. IEEE, 2012.
- [17] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. <http://www.vlfeat.org/>, 2008.
- [18] Lior Wolf, Tal Hassner, Yaniv Taigman, et al. Descriptor based methods in the wild. In *Workshop on Faces in 'Real-Life' Images at ECCV*, 2008.
- [19] Lior Wolf, Tal Hassner, and Yaniv Taigman. Similarity scores based on background samples. In *ACCV*, pages 88–97. Springer, 2010.
- [20] Qi Yin, Xiaoou Tang, and Jian Sun. An associate-predict model for face recognition. In *CVPR*, pages 497–504. IEEE, 2011.