Heterogeneous Face Recognition using Inter-Session Variability Modelling

Tiago de Freitas Pereira, Sébastien Marcel
Idiap Research Institute
hhtp://www.idiap.ch
tiago.pereira@idiap.ch, sebastien.marcel@idiap.ch

Abstract

The task of Heterogeneous Face Recognition consists in to match face images that were sensed in different modalities, such as sketches to photographs, thermal images to photographs or near infrared to photographs. In this preliminary work we introduce a novel and generic approach based on Inter-session Variability Modelling to handle this task. The experimental evaluation conducted with two different image modalities showed an average rank-1 identification rates of 96.93% and 72.39% for the CUHK-CUFS (Sketches) and CASIA NIR-VIS 2.0 (Near infra-red) respectively. This work is totally reproducible and all the source code for this approach is made publicly available.

Face recognition has existed as a field of research for more than 30 years and has been particularly active since the early 1990s. Researchers of many different fields (from psychology, pattern recognition, neuroscience, computer graphics and computer vision) have attempted to create and understand the face recognition task [31].

One of the most challenging tasks in automated face recognition is the matching between face images acquired in heterogeneous environments. Use-cases can cover matching of faces in unconstrained scenarios (e.g. at a distance), with long time lapse between the probe and the gallery and faces sensed in different modalities, such as thermal infrared or near infrared images (NIR) against visible spectra images (VIS). Successful solutions to heterogeneous face recognition can extend the reach of these systems to covert scenarios, such as recognition at a distance or at night-time, or even in situations where no face even exists (forensic sketch recognition).

The key difficulty in matching faces from heterogeneous conditions is that images of the same subject may differ in appearance due to changes in image modality (e.g. between VIS images and NIR images, between VIS images and sketches images) introducing high within class variations. With these variations, a direct comparison between samples generally results in poor matching accuracy [8]. Heterogeneous face recognition algorithms must develop facial representations invariant to these changes.

This work proposes to approach the problem of Hetererogeneous Face Recognition (HFR) as a Session Variability task, modelling the within-class variability using Gaussian Mixture Models (GMM). Experiments carried out with the CASIA NIR-VIS 2.0 Database and CUHK-Face Sketch Database (CUFS) shown competitive results with the current state-of-the-art results. Another contribution of this work is with respect to reproducibility. All the source code used to generate the results and plots are freely available for download. The documentation is done in such way that other researchers are able to reproduce them.

The organization of the paper is the following. In Section 1 we present the prior work for heterogeneous face recognition. In Section 2 the proposed approach is presented in details. In Section 3 the experimental setup and results are presented. Finally in Section 4 the conclusions and future work are presented.

1. Related work

The most frequent heterogeneous face recognition scenarios involve gallery databases with visible light face images (VIS) and probe images from some alternative modality, such as:

- Near infrared (NIR) [8, 12, 9, 5, 7];
- Viewed sketches [8, 23, 24, 7, 20]
- Forensic sketches [8]

A recent study [8] organized the state-of-the-art techniques for heterogeneous face recognition into three approaches:

Synthesis methods: Generates a synthetic version from one modality to another. Once a synthetic version of one modality is generated, the matching can be done with a regular face recognition approaches. In [29], the authors proposed a patch based synthesis in order to synthesize VIS

images to viewed sketches and vice-versa using Multiscale Markov Random Fields. They evaluated the synthetic images using several face recognition algorithms, such as, Eigenfaces [25], Fisherfaces [1], dual space LDA [27] and Random Sampling LDA [28] with a combination of three photo-sketch databases[1] (CUHK, XM2VTS and AR database). In [13], the authors learnt a pixel level mapping between VIS images and viewed sketches with Locally Linear Embeddings (LLE).

Feature-based methods: Feature-based methods encode face images from a pair of image modalities with descriptors that are invariant in both domains. Liao et al. in [12] proposed a method that normalizes both VIS and NIR images using the Tan & Triggs filter [22]. The local descriptor MLBP [18] (with different radii) is extracted from each one of the pre-processed images and after a feature selection step LDA is used to classify each subject. A verification rate of 67.5% was reported under a false acceptance rate of 0.1% on the CASIA-HFB database. Similarly Sifei et al. [14] used a set of different band-pass filters, to "normalize" both VIS and NIR images for posterior recognition. A rank-1 recognition rate of 98.51% was reported. Inspired in gravitational fields to model pixel values, Roy et al. in [20] proposed an illumination invariant feature extractor. The method requires no training model. Experiments carried out with CUHK-CUFS with a biased protocol (see Section 3.4.1) showed a rank-1 recognition rate of 99.96%.

Projection based approaches: The idea of these approaches is to learn a joint mapping that will project images of different nature in a subspace where the image projections can be directly compared. In [8], the authors proposed a generic framework which faces are represented in terms of nonlinear similarities (via kernel function) to a collection of prototype face images from different modalities. The proposed approach, called prototype random subspace (P-RS) was demonstrated on four different heterogeneous scenarios: NIR to VIS, thermal images to VIS, viewed sketch to VIS and forensic sketch to VIS. As VIS to sketch reference results were reported using the CUHK-CUFS database and a Rank-1 of 99% were reported. Finally as a VIS to NIR reference the CASIA HFB was used and a Rank-1 of 98% was reported. In [7] the authors proposed a filter learning approach where the goal is to find the convolutional filter α , where the pixel difference between images from different modalities are the minimum. Experiments with CUHK-CUFSF showed an average Rank-1 of 81.3%.

[1] http://mmlab.ie.cuhk.edu.hk/archive/facesketch.html

2. Proposed approach

As previously mentioned, the key difficulty in heterogeneous face recognition is the high within class variability. To address this task we propose to first model the features from different image modalities with Gaussian Mixture Models ($GMMs$). Then we hypothesize that this variability can be suppressed with a linear shift in the Gaussian Mixture Model (GMM) mean subspace. This approach is called Intersession Variability Modelling (ISV)[26].

2.1. Formulation for heterogeneous face recognition

A GMM is a weighted sum of C multivariate gaussian components:

$$p(o|\Theta_{gmm}) = \sum_{c=1}^{C} w_c \mathcal{N}(o; \mu_c, \Sigma_c), \quad (1)$$

where $\Theta_{gmm} = \{w_c, \mu_c, \sigma_c\}_{\{c=1...C\}}$ are the weights, means and the covariances of the model.

Built on top of GMMs, Intersession Variability Modelling (ISV) proposes to explicitly model the variations between different sessions of the same identity and compensate them during the enrolment and testing time. In our particular task, the term session variability refers to variations regarding to the image modality.

ISV assumes that the session variability is an additive offset (shift) to the GMM mean super-vector space combined with a client specific offset. At **training time** (offline procedure), to model the variability between some hypothetical image modalities A and B, first a GMM is trained with data from different identities. In the literature this GMM is called Universal Background Model (UBM) [19]. The mean super-vector m^{AB} (see Eq. 2) is built by concatenating the means of each gaussian component c of this GMM. Hence, the final super-vector is defined as: $[(\mu_{c=1}^{AB})^T, (\mu_{c=2}^{AB})^T ... (\mu_{c=C}^{AB})^T]$.

Given the j_{th} face sample $\mathcal{O}_{i,j}$ of the identity i, the mean super-vector $\mu_{i,j}$ (independent of the modality) of a GMM can be decomposed as:

$$\mu_{i,j} = m^{AB} + U^{AB} x_{i,j} + D^{AB} z_i, \quad (2)$$

where m^{AB} is the UBM trained with both modalities, U^{AB} is the subspace that contains all possible session effects (also called the within-class variability matrix), $x_{i,j}$ is its associated latent session variable ($x_{i,j} \sim \mathcal{N}(0, I)$), while $D^{AB} z_i$ represents the client offset.

At **enrolment time**, the model for the identity i is obtained by estimating $x_{i,j}$ and z_i using only samples from the modality A. The effect of the session variability for each facial image ($Ux_{i,j}$ in (2)) is then excluded from the final model. In the end, the model of an identity using only samples from modality A is defined as:

$$s_i^A = m^{AB} + D^{AB} z_i \quad (3)$$

.

At **scoring time** (using only samples from modality B), the score is defined as the log-likelihood ratio (LLR) between the target model (estimated only with samples of the modality A) and the UBM (estimated with A and B). Given a set of observations from modality B, $\mathcal{O}^{\mathcal{B}} = \{o_1^B ....o_T^B\}$ claimed to be from the client i, the LLR is defined as follows:

$$h(\mathcal{O}^B|s_i^A) = \sum_{t=1}^{T}[ln(\frac{p(o_t^B|s_i^A + U^{AB}x_{i,j})}{p(o_t^B|m^{AB} + U^{AB}x_{i,j}^{UBM})})] \quad (4)$$

A full derivation on how the U matrix, the latent variable $x_{i,j}$ and the client offset z_i are estimated can be found in [16].

2.2. ISV Intuition for HFR

The Figure 1 shows an intuition on how ISV models heterogeneous data in a toy dataset.

Let's assume that the data points in Figure 1 are our training set. This training set is composed by samples from 2 identities represented by the colors red and blue. The dots in the figure are samples from modality A and the stars are samples from modality B. The UBM (see m in Eq. 2) is then estimated with two Gaussians components (Figure 1 (a),(b),(c) and (d)). The rank of U (Eq. 2) is set to one in order to be plotted in 2D and it is represented by the black arrows (U_1 and U_2).

Let's consider that the green dot in the Figure 1 (b) is one data sample of an unknown identity from modality A that we want to enrol using equation 3. The output super-vector in 3 can be decomposed in terms of each Gaussian component c. This is represented by the cyan diamonds in Figure 1 (b).

Finally for scoring, let's consider that the green star in Figure 1 (c) is one data sample of the same unknown identity, but now from modality B. The magenta diamonds represents the super-vector decomposition with respect to each Gaussian component using this data sample as input. Just for comparison, the red diamonds in the Figure 1 (d) shows the super-vector decomposition using the same sample, but without removing the session factor U^{AB}. It is reasonable to claim that the log-likelihood (see equation 4) obtained in Figure 1 (c) (magenta diamonds) will be higher then the log-likelihood obtained in the Figure 1 (d) (red diamonds). In Figure 1 (c) the cyan and magenta diamonds are almost overlapped. On the other hand, the cyan and red diamonds in Figure 1 (d) are far apart (compared to the magenta diamonds).

It is worth noting that, in this example, only the data is illustrative; the whole model used for this explanation is real. The source code to reproduce these didactically plots is available for download and reproducibility[2].

3. Experiments

This section describes the experimental procedures carried out with two different HFR scenarios: VIS -> NIR and VIS->Sketch. In these two scenarios, VIS images are used to enrol a subject and both NIR or sketches (depending on the database) are used as probes.

All this experimental section is reproducible. The source code to reproduce the experiments with instructions on how to get all plots and tables is released in a python package format[2].

The next subsections explain our experimental setup.

3.1. Databases

This subsection describes the databases used in this work.

3.1.1 CUHK Face Sketch Database (CUFS)

CUHK Face Sketch database[1] (CUFS) is composed by viewed sketches. It includes 188 faces from the Chinese University of Hong Kong (CUHK) student database, 123 faces from the AR database[3] and 295 faces from the XM2VTS database[4].

There are 606 face images in total. For each face image, there is a sketch drawn by an artist based on a photo taken in a frontal pose, under normal lighting condition and with a neutral expression.

There is no evaluation protocol established for this database. Each work that uses this database implements a different way to report the results. In [29] the 606 identities were split in three sets (153 identities for training, 153 for development, 300 for evaluation). The rank-1 identification rate in the evaluation set is used as performance measure. Unfortunately the file names for each set were not distributed.

In [8] the authors created a protocol based on a 5-fold cross validation splitting the 606 identities in two sets with 404 identities for training and 202 for testing. The average rank-1 identification rate is used as performance measure. In [3], the authors evaluated the error rates using only the pairs (VIS $\rightarrow$ Sketch) corresponding to the CUHK Student Database and AR Face Database and in [2] the authors used only the pairs corresponding to the CUHK Stu-

[2] https://pypi.python.org/pypi/bob.paper.CVPRW_2016

[3] http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html

[4] http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/

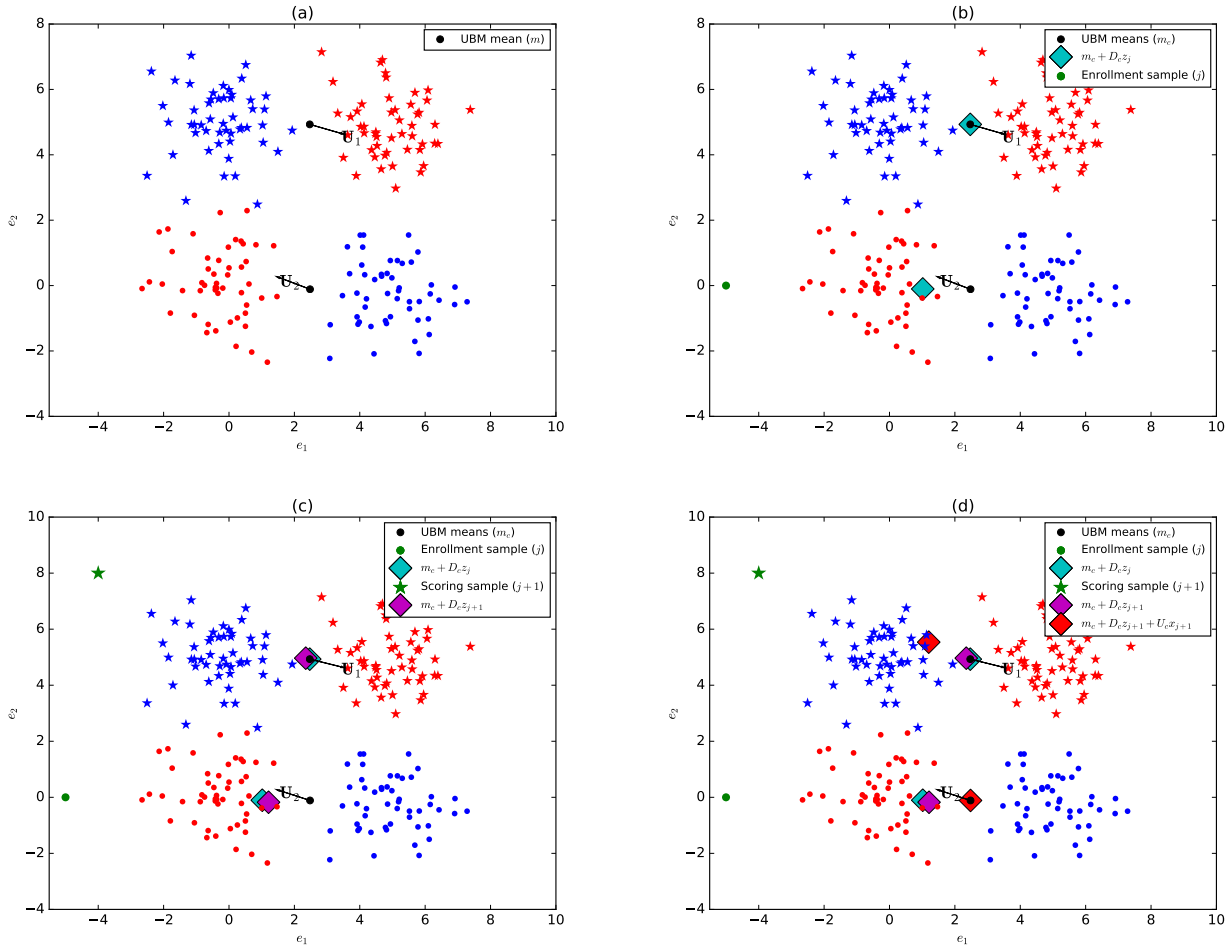

Figure 1. ISV Intuition (a) Estimation of m and U (background model) (b) Enrollment considering the session varibility using a sample j (c) Scoring considering the session varibility for a sample $j+1$(d) Scoring no removing the session varibility $j+1$.

dent Database. In [30] the authors created a protocol based on a 10-fold cross validation splitting the 606 identities in two sets with 306 identities for training and 300 for testing. Also the average rank-1 identification error rate in the test is used to report the results. Finally in [20], since the method does not requires a background model, the whole 606 identities were used for evaluation and also to tune the hype-parameters; which is not a good practice in machine learning. Just by reading what is written in the paper (no source code available), we can claim that the evaluation is biased.

For comparison reasons, we will follow the same strategy as in [8] and do a 5 fold cross-validation splitting the 606 identities in two sets with 404 identities for training and 202 for testing and use the average rank-1 identification rate, in the evaluation set as a metric. For reproducibility purposes, this evaluation protocol is published in a python package format[5]. In this way future researchers will be able to reproduce exactly the same tests with the same identities in each fold (which is not possible today).

3.1.2 CASIA NIR-VIS 2.0 face database

CASIA NIR-VIS 2.0 database [11] offers pairs of mugshot images and their correspondent NIR photos. The images of this database were collected in four recording sessions: 2007 spring, 2009 summer, 2009 fall and 2010 summer, in which the first session is identical to the CASIA HFB database [10]. It consists of 725 subjects in total. There are [1-22] VIS and [5-50] NIR face images per subject. The eyes positions are also distributed with the images. Figure 2 presents some samples of that database.

This database has a well defined protocol and it is pub-

[5] https://pypi.python.org/pypi/bob.db.cuhk_cufs

Figure 2. Samples from CASIA NIR VIS 2.0 Database [11].

licly available for download[6]. We also organized this protocol in the same way as for CUFS database and it is also freely available for download[7]. The average rank-1 identification rate in the evaluation set (called view 2) is used as an evaluation metric.

3.2. Image preprocessing and feature extraction

The goal of this work is to explore the session variability hypothesis for HFR. For simplicity of this analysis the face size and inter-pupil distance were set with constant values. As a reference for those values we used in our experiments the parameters extensively tuned in [6]. This work presents an extensive analysis of face recognition algorithms under different face databases and defined a face size of 80×64 pixels and an inter-pupil distance of 33 pixels, after a geometric normalization, as a good trade-off between face size and recognition rate.

Since the purpose of session variability is to create a background model that handle the gap between different image modalities we will not use any image preprocessing strategy. Any kind of preprocessing in the image level will introduce some noise that is not interesting in our analysis. The analysis of different image preprocessing algorithms under our proposed approach will be discussed in a future work.

Each cropped and geometric normalized face image from each modality is sampled in patches of 12×12 pixels moving the sampled window in one pixel. Then each patch is mean and variance normalized and the first 45 DCT coefficients are extracted. The first coefficient (DC component) is discarded resulting in a feature vector of 44 elements per patch. The feature vectors per patch are not concatenated as in [8]. Each sampled patch is considered as an independent observation.

3.3. ISV Hyper-parameters

The most relevant hyper-parameters for ISV are the number of Gaussian components in m and the rank of U. For both databases we will tune first the number of Gaussian components keeping the rank of $U = 160$. Keeping the number of components that produces the highest rank-1 we will tune the rank of the U.

3.4. Results

This subsection will describe our experiments with the databases presented in the section 3.1.

3.4.1 CUHK Face Sketch Database (CUFS)

Figure 3 (a) presents the CMC plots varying the number of Gaussian components (1024, 512, 256, 128 and 64). The CMC plots represents the averages under the 5 splits with their respective standard deviations. It is possible to observe that there is a correlation between the number of Gaussian components and the average rank-1 identification rate. The highest rank-1 is achieved with 1024 Gaussian components.

Figure 3 (b) presents the CMC plots varying the rank of U (200, 160, 100, 50, 10) keeping the number of Gaussian components to 1024. The highest rank-1 identification rate is achieved with the rank equals to 100.

Table 1 shows the average rank-1 identification rate comparing our proposed approach (ISV) to two references from [8] (P-RS and FaceVACS). Unfortunately, the source code of the approaches from the literature are not available for reproducibility. The best what we can do is to compare with the numbers presented in the paper. Comparing with P-RS, in terms of average rank-1, the difference is 2.1%, which represents ≈ 4 miss classifications. The HFR approach implemented in P-RS is composed by a score a fusion of 180 different face recognition systems (6 systems with 30 bags each). In the approach each face image is geometric normalized with 250×200 pixels keeping an inter-pupil distance of 75 pixels. Three preprocessing strategies is applied: Difference of Gaussian Filter (DoG) [22], Center Surround Divisive Normalization (CSDN) [17] and a Gaussian Filter. For each preprocessed image two different features are extracted: MLBP features [18] (uniform pattern with 59 bins) with 4 different radius (1, 3, 5, 7) and SIFT features [15] (128 features). Compared with our ISV approach, which is composed by only one system instead of

[6] http://www.cbsr.ia.ac.cn/english/NIR-VIS-2.0-Database.html

[7] https://pypi.python.org/pypi/bob.db.cbsr_nir_vis_2

180 complex systems (several bags, different types of feature, different image processing algorithms), the difference of 4 miss classifications doesn't look an enormous gap.

The Table 1 also highlight the rank-1 of a COTS (Commercial Off-The-Shelf) system from FaceVACS[8] that presents presents an average rank-1 of 89.6%, which is lower than the state-of-the-art approaches and ours.

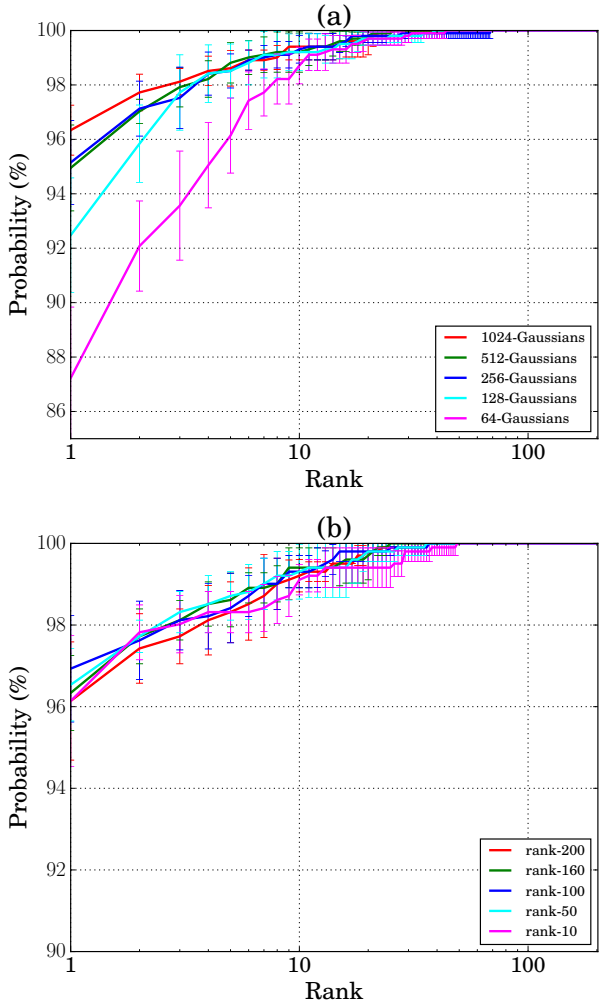

Figure 3. Average CMC plots on the CUHK-CUFS database (a) Varying the number of Gaussian components (1024, 512, 256, 128 and 64) (b) Varying the rank of U (200, 160, 100, 50 and 10) keeping $m = 1024$.

3.4.2 CASIA NIR-VIS 2.0 face database

Figure 4 (a) presents the CMC plots varying the number of Gaussian components (1024, 512, 256, 128 and 64). The CMC plots represents the averages under the 5 splits with their respective standard deviations. It is possible to observe the same trend as in CUHK-CUFS and the 1024 Gaussian components presents the highest rank-1 identification rate.

Figure 4 (b) presents the CMC plots varying the rank of U (200, 160, 100, 50, 10) keeping the number of Gaussian components to 1024. The highest rank-1 identification rate is achieved with the rank equals to 200.

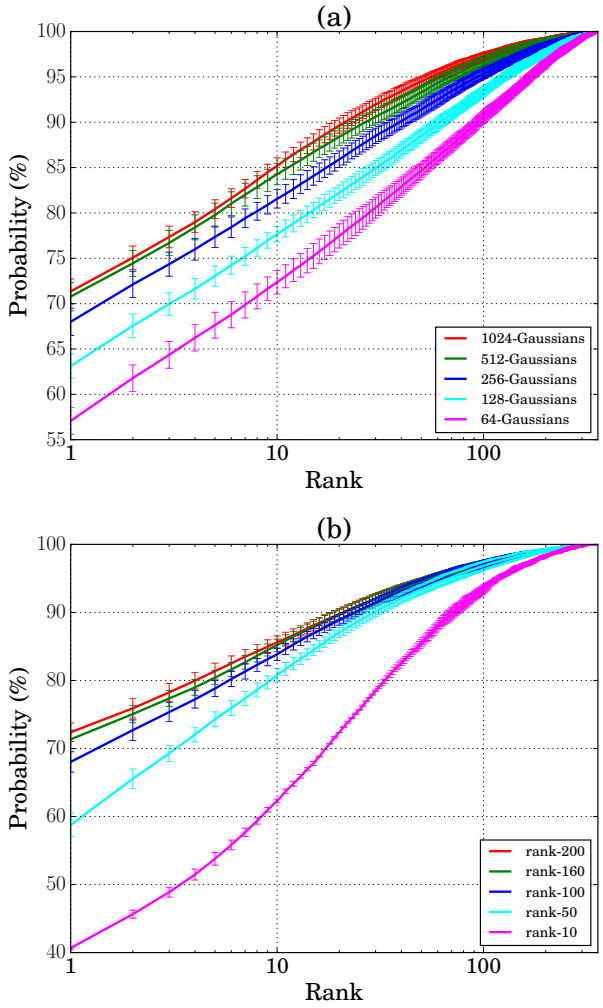

Figure 4. CMC plots on the CASIA NIR-VIS 2.0 database (a) Varying the number of Gaussian Components (1024, 512, 256, 128 and 64) (b) Varying the rank of U (200, 160, 100, 50 and 10) keeping $m = 1024$.

Table 2 shows the rank-1 identification rate compared with the state of the art approaches. As in section 3.4.1, the source code of the approaches from the literature are not available for reproducibility. The best what we can do is to compare with the numbers presented in the paper.

We can observe that the best configuration of our ISV approach is far better than the proposed baseline. It presents

[8] http://www.cognitec.com/facevacs-videoscan.html

Table 1. Average Rank 1 one recognition rate under 5 splits of the proposed approach (ISV: $m = 1024$ and $rank(U) = 100$)

Method	Mean accuracy	Std. Deviation
P-RS as in [8] (section 7.2)	99.%	not informed
Face VACS in [8] (section 7.2)	89.6%	not informed
ISV	**96.9%**	**1.3%**

Table 2. Average rank 1 one recognition rate on View2 under 10 splits of the proposed approach (ISV: $m = 1024$ and $rank(U) = 200$)

Method	Mean accuracy	Std. Deviation
Original baseline [11] (Table 2)	23.70%	1.89%
CDFL in [7] (Table I)	71.5%	1.4%
CMFL in [21] (Table VII)	43.8%	not informed
DSIFT in [4] (Table II)	73.28%	1.10%
FaceVACS in [4] (Table I)	58.56%	1.19%
ISV	**72.39%**	**1.35%**

an average rank-1 identification rate of 72.39% compared with 23.70%. Comparing it with the DSIFT, in terms of average rank-1 identification rate, they are $\approx 1\%$ better (73.28% against 72.39%).

As for the CUFS database, Table 2 presents a comparison with a COTS system from Face VACS. In terms of rank-1 identification rate, our ISV approach (72.39%) is far better than the COTS (58.56%).

It is worth noting that, unlike other techniques, we did not use any image preprocessing strategy. There is still a window of improvement left for future work.

4. Conclusion

This preliminary work investigates the task of HFR as session variability problem. ISV showed competitive results in two different image modalities. Experiments with CUFS showed an average rank-1 identification rate of 96.93%. With CASIA NIR-VIS 2.0 an average rank-1 identification rate of 72.39% was achieved.

This work focused on the proposal and application of session variability for HFR. Unlike techniques from the literature, no image preprocessing was used so far in our study. A study on how different image processing techniques impacts in our proposed approach as well as evaluations with other HFR databases with different image modalities will be covered in future work.

Unlike other studies from literature all the source code used in this work as well as execution instructions are freely available for reproducibility purposes. This is an important contribution of this work.

5. Acknowledgment

The development of this work has received funding from the Swiss National Science Foundation (SNSF) under the HFACE project and from the European Community's Seventh Framework Programme (FP7) under grant agreement 284989 (BEAT) and the Swiss Center for Biometrics Research and Testing.

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997. 2

[2] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. On matching sketches with digital face images. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1–7. IEEE, 2010. 3

[3] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. Memetically optimized mcwld for matching sketches with digital face images. *Information Forensics and Security, IEEE Transactions on*, 7(5):1522–1535, 2012. 3

[4] T. I. Dhamecha, P. Sharma, R. Singh, and M. Vatsa. On effectiveness of histogram of oriented gradient features for visible to near infrared face matching. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 1788–1793, Aug 2014. 7

[5] D. Goswami, C.-H. Chan, D. Windridge, and J. Kittler. Evaluation of face recognition system in heterogeneous environments (visible vs nir). In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 2160–2167, 2011. 1

[6] M. Günther, L. El Shafey, and S. Marcel. Face recognition in challenging environments: An experimental and reproducible research survey. Feb. 2016. 5

[7] Y. Jin, J. Lu, and Q. Ruan. Coupled discriminative feature learning for heterogeneous face recognition. *Information Forensics and Security, IEEE Transactions on*, 10(3):640–652, 2015. 1, 2, 7

[8] B. F. Klare and A. K. Jain. Heterogeneous face recognition using kernel prototype similarities. *Pattern Analysis and Ma-*

chine Intelligence, IEEE Transactions on, 35(6):1410–1422, 2013. 1, 2, 3, 4, 5, 7

[9] S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):627–639, 2007. 1

[10] S. Z. Li, Z. Lei, and M. Ao. The hfb face database for heterogeneous face biometrics research. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2009. 4

[11] S. Z. Li, D. Yi, Z. Lei, and S. Liao. The casia nir-vis 2.0 face database. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 348–353. IEEE, 2013. 4, 5, 7

[12] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li. Heterogeneous face recognition from local structures of normalized appearance. pages 209–218, 2009. 1, 2

[13] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma. A nonlinear approach for face sketch synthesis and recognition. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 1005–1010 vol. 1, 2005. 2

[14] S. Liu, D. Yi, Z. Lei, and S. Z. Li. Heterogeneous face image matching using multi-scale features. In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 79–84. IEEE, 2012. 2

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 5

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

[17] E. Meyers and L. Wolf. Using biologically inspired features for face processing. *International Journal of Computer Vision*, 76(1):93–104, 2008. 5

[18] M. Pietikäinen. *Computer vision using local binary patterns*, volume 40. Springer, 2011. 2, 5

[19] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital signal processing*, 10(1):19–41, 2000. 2

[20] H. Roy and D. Bhattacharjee. Local-gravity-face (lg-face) for illumination-invariant and heterogeneous face recognition. *IEEE Transactions on Information Forensics and Security*, PP(99):1–1, 2016. 1, 2, 4

[21] M. Shao and Y. Fu. Cross-modality feature learning through generic hierarchical hyperlingual-words. *IEEE Transactions on Neural Networks and Learning Systems*, PP(99):1–13, 2016. 7

[22] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *Analysis and Modeling of Faces and Gestures*, pages 168–182. Springer, 2007. 2, 5

[23] X. Tang and X. Wang. Face sketch synthesis and recognition. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 687–694. IEEE, 2003. 1

[24] X. Tang and X. Wang. Face sketch recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):50–57, 2004. 1

[25] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991. 2

[26] R. Vogt and S. Sridharan. Explicit modelling of session variability for speaker verification. *Computer Speech & Language*, 22(1):17–38, 2008. 2

[27] X. Wang and X. Tang. Dual-space linear discriminant analysis for face recognition. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–564. IEEE, 2004. 2

[28] X. Wang and X. Tang. Random sampling lda for face recognition. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–259. IEEE, 2004. 2

[29] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(11):1955–1967, 2009. 1, 3

[30] D. Yi, Z. Lei, and S. Z. Li. Shared representation learning for heterogenous face recognition. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, volume 1, pages 1–7. IEEE, 2015. 4

[31] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–459, 2003. 1