Scalable Probabilistic Models: Applied to Face Identification in the Wild

Laurent El Shafey*, Sébastien Marcel

Biometric Person Recognition Group
Idiap Research Institute
Rue Marconi 19 PO Box 592
1920 Martigny
laurent.el-shafey@idiap.ch
sebastien.marcel@idiap.ch

Abstract: Face recognition is a mature field in the biometric community. While existing systems perform well under controlled recording conditions, mismatch caused by the use of different sensors or a lack of cooperation from the subject still significantly affects performance, especially in challenging scenarios such as in forensics. In this paper, we propose to address this problem of mismatch using probabilistic approaches that model the variability of face images. Specifically, a face identification evaluation is conducted on the challenging Labeled Faces in the Wild database, these models leading to a relative improvement in accuracy of about 25% when compared to the NIST CSU baselines.

1 Introduction

Automatic face recognition is the task of recognizing people from digital face images [JL05]. Early studies were conducted in the sixties already and it is now a mature topic in the biometrics field [ZCPR03]. These decades of intensive research have led to the development of systems that work well under controlled laboratory conditions. However, they are still severely impacted under the wildly varying conditions in terms of pose, illumination or expression encountered in many real world scenarios.

A large number of applications are possible ranging from access control and surveillance, to data management and personalization. For instance, border control based on face recognition technology has been successfully deployed in few airports [SHG12]. In such a scenario, subjects are cooperating with the system, which restricts the amount of variability. Another interesting application is surveillance motivated by the growth of closed-circuit television (CCTV) systems. This is of particular interest in forensic science [MV12], since

*Acknowledgments: Part of this research work has been carried out while working at the Netherlands Forensic Institute (NFI). I am hence grateful to Dr. Didier Meuwly and Arnout Ruifrok. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7) under grant agreements 238803 (BBfor2) and 284989 (BEAT).

face recognition technology can then be employed to reduce the quantity of information to be processed manually, when criminal or terrorist investigations are performed [KJ13]. In this case, the lack of cooperation from the subjects leads to a large intra-class variability of face images.

In this study, we explore the use of probabilistic models such as inter-session variability modeling (ISV) and probabilistic linear discriminant analysis (PLDA) to address this problem of face identification in uncontrolled environments. These approaches both models the data as well as the within-class variability of the data to improve the accuracy of the classification in such challenging conditions. In addition, these techniques need to process increasingly large amounts of data and scalability is, hence, a key aspect for such real world scenarios. The experimental evaluation is conducted on the Labeled Faces in the Wild (LFW) database using a well defined identification protocol. A more detailed version of this study can be found in [ES14].

The remainder of this paper is structured as follows. First, the probabilistic models are succinctly presented (Section 2). Next, the experimental setup is described and identification results are discussed (Section 3), before concluding this paper (Section 4).

2 Probabilistic Models

Several approaches for face recognition have been proposed during the past two decades, and we refer to [ZCPR03] for an overview. In this study, we investigate the use of a specific subset of techniques, probabilistic latent variable models, to address the problem of unconstrained face identification. A succinct description is provided in this section.

2.1 Gaussian Mixture Models (GMM)

The idea of this approach consists of modeling parts-based features with a Gaussian mixture model distribution [LC04] following the technique for speaker recognition given in [RQD00]. These parts-based features are obtained by decomposing the face image into a set of overlapping blocks and extracting a descriptor from each block.

After feature extraction, the distribution of the resulting feature vectors for each client is modeled by a GMM, estimated using background model adaptation [RQD00]. Background model adaptation utilizes a universal background model (UBM) as a prior for deriving client models using maximum *a posteriori* (MAP) adaptation. In practice, we only adapt the means of the GMM components, which has been shown to be effective. Using the GMM supervector notation, MAP adaptation can be written:

$$\boldsymbol{g}_i = \boldsymbol{m} + \boldsymbol{d}_i, \tag{1}$$

where $\boldsymbol{d}_i$ is a client-specific offset, and $\boldsymbol{g}_i$ and $\boldsymbol{m}$ are the mean supervectors of the client and the UBM, respectively.

The feature vectors $\boldsymbol{o}_t^k$ from a test sample $\mathcal{O}_t$, can then be compared to a client-specific model, by producing a log-likelihood ratio (LLR) as follows:

$$s_{\mathbf{GMM}}\left(\boldsymbol{\mathcal{O}}_t, \boldsymbol{g}_i\right) = \sum_{k=1}^{K} \left(\log(p(\boldsymbol{o}_t^k \mid \boldsymbol{g}_i)) - \log(p(\boldsymbol{o}_t^k \mid \boldsymbol{m}))\right). \tag{2}$$

2.2 Variability Compensation using GMMs

When enrolling a client-specific GMM, the resulting model should ideally be robust to any variations within the client's enrollment samples. However, this variation is not accounted for in GMM MAP adaptation, and, hence, it may yield a suboptimal model, particularly in the case of limited enrollment data.

Several techniques built on top of GMMs have been proposed to improve the robustness to such variabilities. They assume that the particular conditions of a session result in an additive offset to the GMM mean supervector:

$$\boldsymbol{\mu}_{i,j} = \boldsymbol{g}_i + \boldsymbol{u}_{i,j}, \tag{3}$$

where $\boldsymbol{u}_{i,j}$ is the session-dependent offset for the j'th image of client i, and $\boldsymbol{\mu}_{i,j}$ is the resulting mean supervector of the GMM that best represents the image ($\boldsymbol{\mathcal{O}}_{i,j}$).

Inter-session variability (ISV) modeling and joint factor analysis (JFA) are two such approaches. They assume that the within-client variation is contained in a linear subspace of the GMM mean supervector space. That is:

$$\boldsymbol{u}_{i,j} = \boldsymbol{U}\boldsymbol{x}_{i,j}, \tag{4}$$

where $\boldsymbol{U}$ is the low-dimensional subspace that contains within-client variation, and $\boldsymbol{x}_{i,j} \sim \mathcal{N}\left(0, \boldsymbol{I}\right)$. The ISV and JFA approaches differ in the way the client-specific term $\boldsymbol{g}_i$ is learned.

Another related approach is total variability modeling (TV), which aims to extract low-dimensional factors $\boldsymbol{w}_{i,j}$, so-called *i-vectors*, from biometric samples $\boldsymbol{\mathcal{O}}_{i,j}$. More formally, the TV approach can be described in the GMM mean supervector space by:

$$\boldsymbol{\mu}_{i,j} = \boldsymbol{m} + \boldsymbol{T}\boldsymbol{w}_{i,j}, \tag{5}$$

where $\boldsymbol{T}$ is the low-dimensional total variability subspace and $\boldsymbol{w}_{i,j}$ the low-dimensional i-vector, which is assumed to follow a normal distribution $\mathcal{N}\left(0, \boldsymbol{I}\right)$.

These approaches and their corresponding (approximate) scalable training procedures using expectation-maximization (EM) are described in more details in [ES14].

2.3 Probabilistic Linear Discriminant Analysis (PLDA)

Probabilistic linear discriminant analysis (PLDA) is a probabilistic framework that incorporates both between-class and within-class information. In contrast to the previous

GMM-based approaches, a single training step aims at modeling both the features and the variability of these features. More formally, PLDA assumes that the j-th sample of client i is generated by:

$$\boldsymbol{o}_{i,j} = \boldsymbol{F}\boldsymbol{h}_i + \boldsymbol{G}\boldsymbol{k}_{i,j} + \boldsymbol{\epsilon}_{i,j} \,, \tag{6}$$

where $\boldsymbol{F}$ and $\boldsymbol{G}$ are the subspaces describing the between-class and within-class variations, respectively, $\boldsymbol{h}_i$ and $\boldsymbol{k}_{i,j}$ are the associated latent variables, which are assumed to be normally distributed $\mathcal{N}(0, \boldsymbol{I})$, and $\boldsymbol{\epsilon}_{i,j}$ represents the residual noise, which is supposed to follow a Gaussian distribution $\mathcal{N}(0, \boldsymbol{\Sigma}_\epsilon)$.

The parameters $\boldsymbol{\Theta}_{\text{plda}} = \{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{\Sigma}_\epsilon\}$ of this model are learned using an EM algorithm over a training set of features. Once the model has been trained, the feature vector $\boldsymbol{o}_t$ extracted from a probe sample $\boldsymbol{\mathcal{O}}_t$ is compared to the feature vector $\boldsymbol{o}_i$ representing a client i by computing the LLR score:

$$s_{\textbf{PLDA}}\left(\boldsymbol{o}_t, \boldsymbol{o}_i\right) = \log\left(\frac{p(\boldsymbol{o}_t, \boldsymbol{o}_i \mid \boldsymbol{\Theta})}{p(\boldsymbol{o}_t \mid \boldsymbol{\Theta})p(\boldsymbol{o}_i \mid \boldsymbol{\Theta})}\right) \,. \tag{7}$$

Here, $p(\boldsymbol{o}_t, \boldsymbol{o}_i \mid \boldsymbol{\Theta})$ is the likelihood that the features $\boldsymbol{o}_t$ and $\boldsymbol{o}_i$ share the same latent identity variable $\boldsymbol{h}_i$ and, hence, are coming from the same client, whereas $p(\boldsymbol{o}_t \mid \boldsymbol{\Theta})p(\boldsymbol{o}_i \mid \boldsymbol{\Theta})$ is the likelihood that the features $\boldsymbol{o}_t$ and $\boldsymbol{o}_i$ have different latent identity variables $\boldsymbol{h}_t$ and $\boldsymbol{h}_i$ and, therefore, are from different clients.

At both training and test time, a critical aspect of this model is the scalability with respect to the number of samples per class. This has been addressed in [ES14].

3 Experimental Evaluation

In this section, we evaluate the accuracy of the systems considering an identification scenario on the Labeled Faces in the Wild (LFW) database [HRBLM07]. Source code required to reproduce the experiments is available online[1] and relies on the Bob library.[2]

3.1 Identification using Labeled Faces in the Wild

Several face databases have been acquired over the years, such that researchers are able to benchmark their algorithms on them. They have different sizes, both in terms of number of identities and number of samples, and different properties, such as controlled or uncontrolled recording conditions. The LFW database is a good candidate for face identification in the context of forensic investigation, since it contains a reasonably large number of identities and images have large variabilities. However, the original evaluation protocol on LFW is a verification one. Therefore, we define a new identification protocol by splitting the data into three sets for training, development and evaluation purposes. The training set

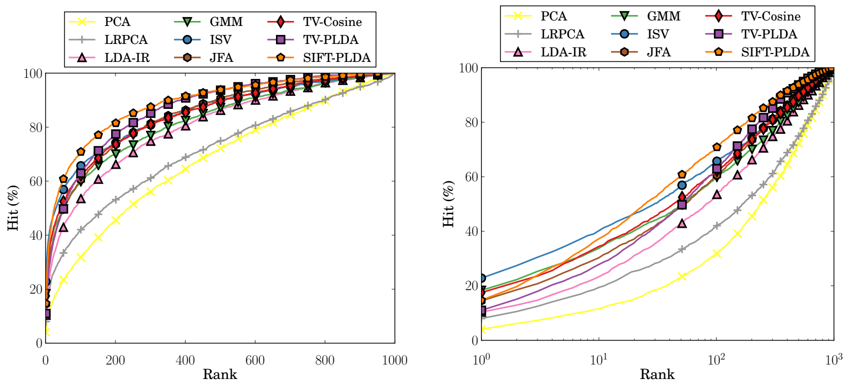
[1] `http://pypi.python.org/pypi/xbob.thesis.elshafey2014`

[2] `http://www.idiap.ch/software/bob`

consists of 280 identities, which all have at least 10 samples, to be able to accurately model within-class variations. The remaining identities were split as follows: 400 for the development set and $1,000$ for the evaluation set. Model enrollment is performed using the first image labeled as 0001 of each client. This finally results in a challenging identification protocol ($1,000$-class problem) using images of faces in the wild.

3.2 Experimental Results

We evaluate a set of nine face recognition systems. Two baselines, **LRPCA** and **LDA-IR**, are taken from the *CSU Face Recognition Resources*,[3] which provide the baseline algorithms for *the Good, the Bad & the Ugly* (GBU) challenge [PBD+11] and we also provide the simple Eigenface one (**PCA**). The other systems, **GMM**, **ISV**, **JFA**, **TV-Cosine**, **TV-PLDA** and **SIFT-PLDA** rely on the probabilistic approaches introduced in Section 2. **TV-Cosine** and **TV-PLDA** use the same TV extraction approach, but employ different modeling strategies. In contrast, **TV-PLDA** and **SIFT-PLDA** rely on the same PLDA modeling approach, but use different features. All the hyperparameters of these systems are tuned on the development set of the LFW database, while cumulative match curves (CMC) on the evaluation set are shown on Figure 1.



(a) Identification on LFW (rank on a linear scale) (b) Identification on LFW (rank on a logarithm scale)

Figure 1: IDENTIFICATION PERFORMANCE OF THE SYSTEMS ON LFW.

Interestingly, the best performing system depends on the operating point. Considering an identification rank in the range $[1, 20]$, **ISV** outperforms other systems, the recognition rate at rank 1 being slightly above 20%. If the system can afford to return a larger output list of potential matches (more than 20), **SIFT-PLDA** is then the most accurate system, and for even larger list, **TV-PLDA** becomes a good choice as well. These results suggest that

[3] http://www.cs.colostate.edu/facerec/algorithms/baselines2011.php

these techniques are able to cope with the session variability issue up to a certain level. Specifically, when compared to the best baseline (**LRPCA**), a relative improvement in accuracy of around 25% at rank 200 is observed with **SIFT-PLDA**.

4 Conclusions

In this paper, we addressed the problem of face identification in uncontrolled conditions. For this purpose, we employed probabilistic approaches that model the variability of face images. The conducted evaluation on the Labeled Faces in the Wild database showed that these techniques are pretty accurate when compared to the NIST/CSU baselines, a relative improvement in accuracy of around 25% being observed at rank 200. Future work may consider approaches to compensate for specific variations such as pose, which would help to reach a higher level of performance required by forensic investigation applications.

References

[ES14] L. El Shafey. *Scalable Probabilistic Models for Face and Speaker Recognition*. PhD thesis, École Polytechnique Fédérale de Lausanne (EPFL), April 2014.

[HRBLM07] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[JL05] A. K. Jain and S. Z. Li. *Handbook of Face Recognition*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[KJ13] J.C. Klontz and A.K. Jain. A case study of automated face recognition: The Boston marathon bombings suspects. *Computer*, 46(11):91–94, November 2013.

[LC04] S. Lucey and Tsuhan Chen. A GMM parts based face representation for improved verification through relevance adaptation. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages II–855–II–861 Vol.2, June 2004.

[MV12] D. Meuwly and R. N. J. Veldhuis. Forensic biometrics: From two communities to one discipline. In *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG)*, September 2012.

[PBD+11] P. J. Phillips, J. R. Beveridge, B. A. Draper, et al. An introduction to the Good, the Bad, & the Ugly face recognition challenge problem. In *IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG)*, 2011.

[RQD00] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 10:19–41, 2000.

[SHG12] L.J. Spreeuwers, A.J. Hendrikse, and K. J. Gerritsen. Evaluation of automatic face recognition for automatic border control on actual data recorded of travellers at Schiphol Airport. In *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG)*, September 2012.

[ZCPR03] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, December 2003.