PRIVACY AND DATA PROTECTION IN BIOMETRICS

Author: Daniel Pérez Cabo

Advisors: Fernando Pérez González
Daniel González Jiménez

Universidad de Vigo

gradiant
Connectivity · Intelligence · Security
for your business

Motivation of the work

Biometrics have emerged as a more secure alternative to passwords and their use is exponentially increasing.

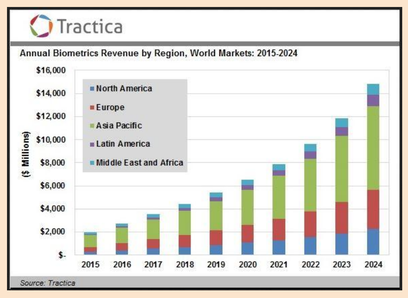


Figure 1. Mobile market estimated revenue for next years

However it is necessary to keep in mind that biometrics may have weaknesses related to users' privacy and security:

- Is it possible for users to cancel and renew their biometric templates?
- Could an attacker be able to infer the biometric sample from the biometric template?
- Could an attacker be able to link a given user through different services?

Biometric systems must handle these issues in order to preserve users' privacy.

Thesis objectives

The main objectives of this thesis are as follows:

1. Understand the fundamentals of biometrics and data protection schemes.
2. Develop a face-based biometric system which achieves state of the art performance and, at the same time, preserves users' privacy. Therefore this objective will be divided in two:
   a. A biometric recognition module with state of the art performance in realistic, large databases.
   b. A privacy module which preserves biometrics performance and meets the privacy characteristics specified at ISO/IEC 24745:
      - Unlinkability
      - Revokability
      - Irreversibility

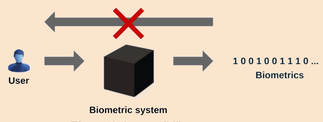
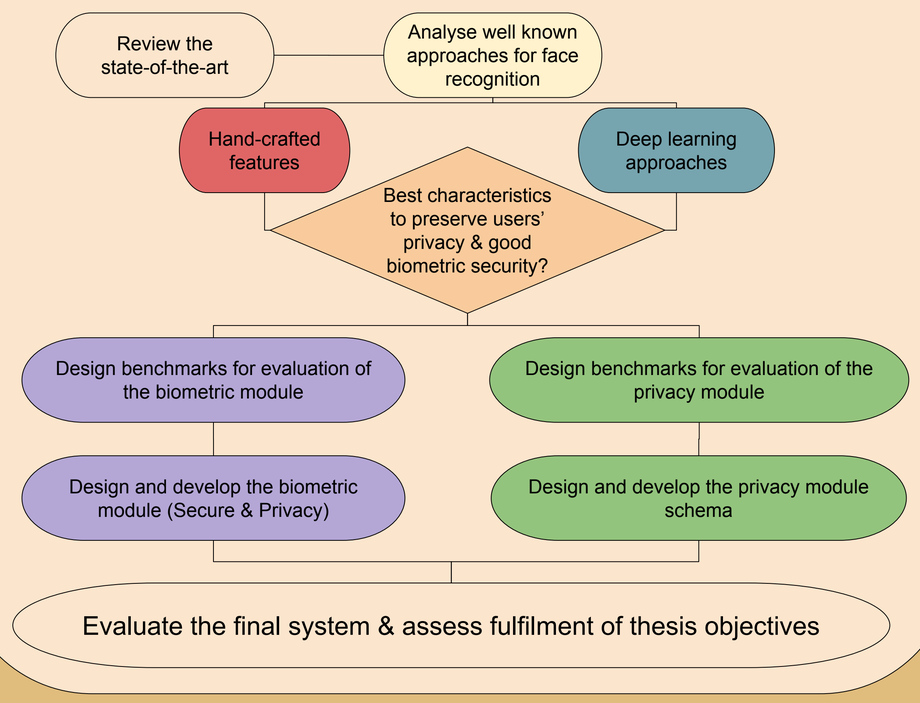
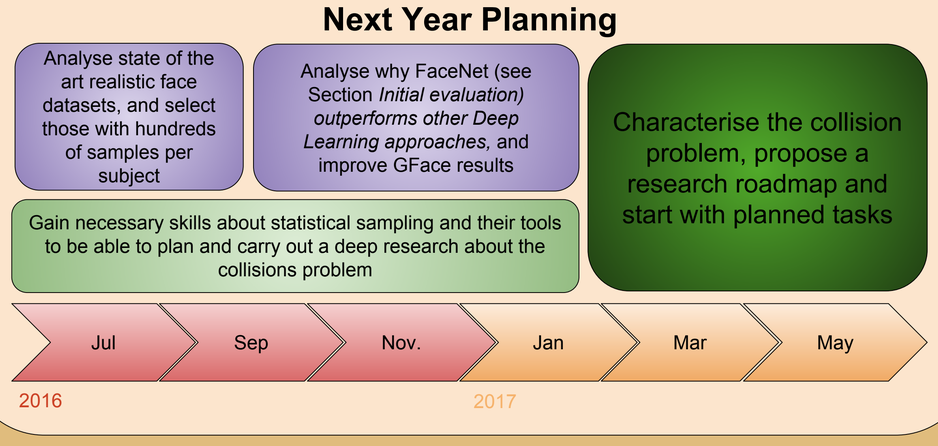


Figure 2. Irreversibility

Research Plan



Next Year Planning



Results And Discussions

Databases

To perform **the training** of the initial system, **VGG Face dataset** [1] was selected since it contains more than **2.000 identities with 1.000 images per identity** approximately. All of them labeled with the bounding box of the faces.

The **LFW database** was selected as the baseline for **evaluating** the ability to discriminate between users due to the following properties:

- This database contains **more than 5,000 identities**.
- It does **not contain any identity of the training dataset**.
- De facto benchmark allowing for fair comparison against state-of-the-art approaches.

But the LFW database is not the only one database that will have to be used since it **has only a few images per identity**. Thus, it does not allow to perform the intra-class variation analysis and another database should be selected.

Designed biometric module

The **convolutional neural network designed** along this year, named ***GFace*** (see Figure 3), is formed by **nine convolutional** layers of which three of them are grouped in a NIN layer [3] and **three fully-connected** layers.

To train GFace network the **Stochastic Gradient Descent** [4] with **Momentum** [5] method was used and the net's weights were initialized using **Xavier Initialization** as described in [6].

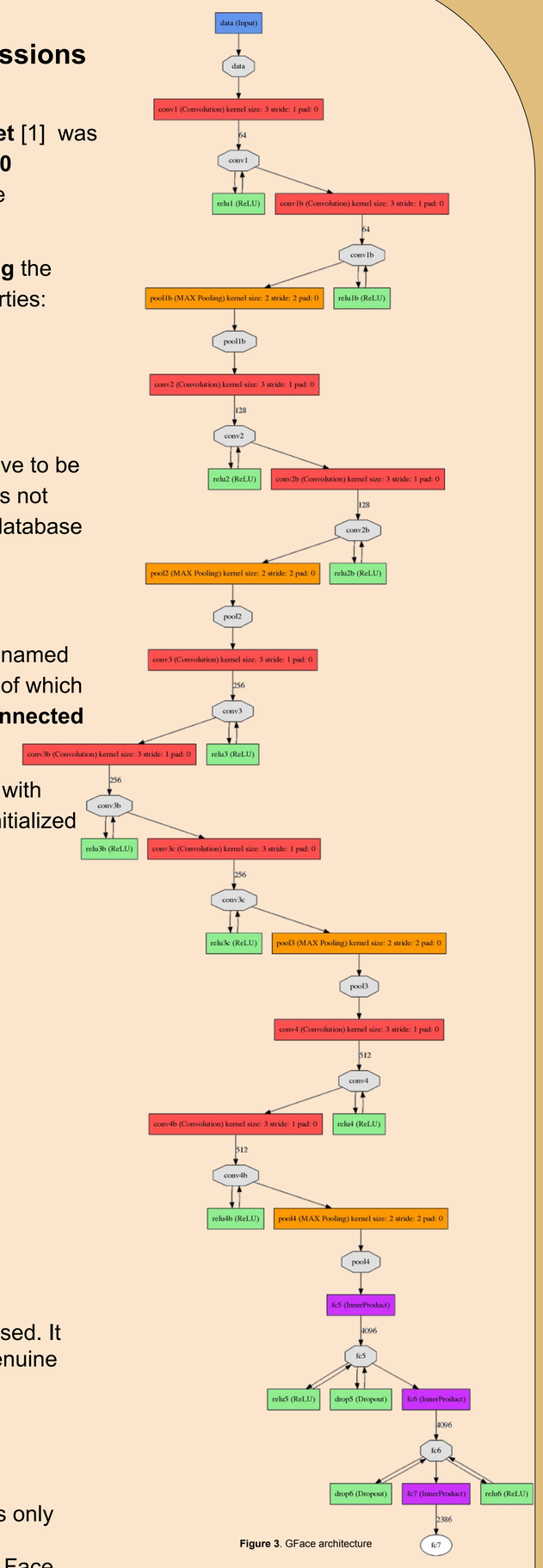


Figure 3. GFace architecture

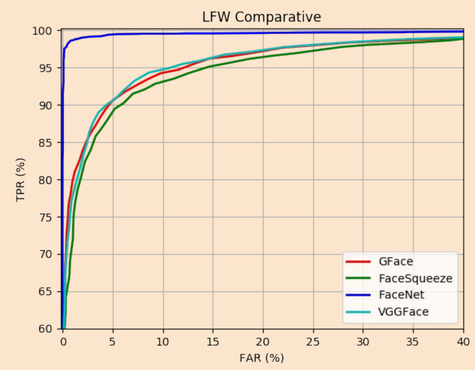


Figure 4. LFW performance comparison between GFace, FaceSqueeze FaceNet and VGGFace networks

Initial evaluation

In order to carry out the initial evaluation, LFW "View 2" was used. It consists of 6000 pairs of images divided in 10 folds with 300 genuine pairs and 300 impostor pairs. The validation is carried out by cross-validation.

The evaluation has also been carried out for three alternative DNN-based face recognition systems:

- **FaceSqueezeNet** [7]: Light system with convolutional layers only (same training as GFace).
- **VGGFace** [8]: Very Deep Neural Network trained with VGG Face dataset.
- **FaceNet** [9]: "Face Embeddings" oriented training with inception layers.

The system TPR (true positive rate) for EER (equal error rate) are shown in Table 1 while ROC curves are shown in Figure 4.

System	TruePositiveRate(EER)
GFace	92,63 % +- 1,38
FaceSqueezeNet	91,85 % +- 1,12
FaceNet	98,73 % +- 0,40
VGGFace	93,03 % +- 1,26

Table 1. True positive rate at equal error rate point for LFW View 2

Collision analysis

The main idea is being able to **estimate** the **collision probability** between two any users. To do that, it will be necessary to:

- **Model** the **intra-class** feature vectors **distribution**.
- **Model** the users' **distribution along the feature space**.

In order to show an example, an ideal scenario is plotted in Figure 5 where two users have null collision probability.

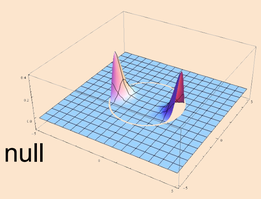


Figure 5. Ideal scenario with two dimensional feature vector for two users with null collision probability

Conclusions

As the initial evaluation shows, FaceNet offers very good performance but, what does cause this improvement?

- The complexity of the network?
- The embeddings approach?
- The training data?

Next year, finding the answer to this question will be investigated to improve the system performance.

Related to the **collision analysis**, it is a key point which **could conduct to the solution** of this research work so next year, the greatest effort will focus on this topic.

References

[1] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments". University of Massachusetts, Amherst, Technical Report 07-49, October, 2007
[2] O. M. Parkhi, A. Vedaldi, A. Zisserman. "Deep Face Recognition". British Machine Vision Conference, 2015
[3] Lin, M., Chen, Q., and Yan, S. "Network in network". In Proc. ICLR, 2014
[4]L. Bottou. Stochastic Gradient Descent Tricks. Neural Networks: Tricks of the Trade: Springer, 2012
[5]Polyak, B.T. "Some methods of speeding up the convergence of iteration methods". USSR Computational Mathematics and Mathematical Physics, 4(5):1–17, 1964
[6] Glorot, X. and Bengio, "Y. Understanding the difficulty of training deep feedforward neural networks". In Proc. AISTATS, volume 9, pp. 249–256, 2010
[7] Forrest N. Iandola,Matthew W. Moskewicz, Khalid Ashraf, Song Han, William J.Dally & Kurt Keutzer (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and >1MB. CoRR, abs/1602.07360
[8] O. M. Parkhi, A. Vedaldi, A. Zisserman. "Deep Face Recognition". British Machine Vision Conference, 2015
[9] Florian Schroff,Dmitry Kalenichenko & James Philbin (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. CoRR, abs/1503.03832