Master in Computer Vision Barcelona

Master Thesis Dissertation

Comparing classical and deep approaches for face recognition in a smartgym application

Author: Gonzalo Benito
gonzabenito@gmail.com

Advisors: Sergio Escalera, UB and CVC
Josep Lladós, UAB and CVC

Organization

1. Context and motivation
2. Objectives
3. State of the Art
4. Methods
5. Experimental setup
6. Results
7. Conclusions and future work
8. Appendix

Motivation

1. **Football Club Barcelona's smartgym** project aims to develop a system to aid **gymnasium training**.
2. System should **identify and track down athletes** within the facility, allowing self configuration of the exercise machines according to the user and computing training times of each one of them.
3. The **face recognition** method should perform with precision in real-time on reduced sets of subjects and samples per individual.

Objectives

1. Compare **three different** approaches to the **face recognition** task, 2 classical and one deep learning based.
2. Evaluate **precision performance** over different **samples per subject** size.
3. Compute face recognition **times**.
4. Analyze viability for **experimental field implementation** and further tests.

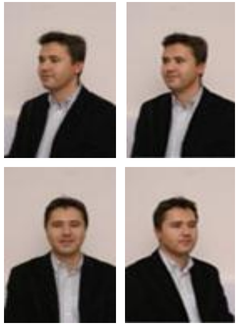
State of the Art

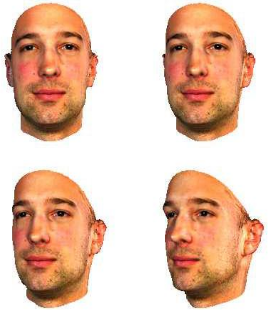
Training data is crucial for the final algorithms **performance**. Many current methods rely on really **big databases**, most of them being still **proprietary**.

Public databases provide a platform for the study of possible solutions for the project.

State of the Art

Face recognition Databases examples





[1] FERET:
1. 14126 images.
2. 1199 subjects.
3. ~12 samples per subject.
4. 180° face rotation.
5. **First** big face database.

[2] SCface:
1. 4160 images.
2. 130 subjects.
3. 32 samples per subject.
4. 180° face rotation.
5. Samples with **surveillance cameras**

[3] MIT-CBCL:
1. 3240 images.
2. 10 subjects.
3. 324 samples per subject.
4. 0° to 34° face rotation.
5. **3D synthetic masks.**

[1] - P. J. Phillips, S. Z. Der, P. J. Rauss, and O. Z. Der, FERET (face recognition technology) recognition algorithm development and test results. Army Research Laboratory Adelphi, MD, 1996.
[2] - M. Grgic, K. Delac, and S. Grgic, "Scface – surveillance cameras face database," Multimedia Tools and Applications, vol. 51, no. 3, pp. 863–879, 2011.
[3] - B. Heisele, B. Weyrauch, V. Blanz, and J. Huang, "Component-based face recognition with 3d morphable models," 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, vol. 05, p. 85, 2004.

State of the Art

Face recognition method examples

Name	Method	Images (millions)	Accuracy	Code
FaceNet [1]	CNN	200	99.63%	Unofficial
DeepID3 [2]	CNN	0.29	99.53%	No
VGG Face EL [3]	CNN	5	99.13%	Yes
DeepFace [4]	CNN	4.4	97.35%	Unofficial
TL Joint Bayesian [5]	Joint Bayesian	0.099	96.33%	No
High-dim LBP [6]	LBP	0.099	95.17%	Unofficial

Top Face recognition methods tested on Labeled Faces in the Wild.

[1] - F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering." in CVPR. IEEE Computer Society, 2015.
[2] - Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks." CoRR, vol. abs/1502.00873, 2015.
[3] - O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition." in BMVC, X. Xie, M. W. Jones, and G. K. L. Tam, Eds. BMVA Press, 2015 .
[4] - Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Jun. 2014.
[5] - X. Cao, D. P. Wipf, F. Wen, G. Duan, and J. Sun, "A practical transfer learning algorithm for face verification." in ICCV. IEEE Computer Society, 2013.
[6] - D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: A joint formulation." in ECCV (3), ser. Lecture Notes in Computer Science, A. W. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds., vol. 7574. Springer, 2012.

Methods

The three methods tested are:

1. Eigenfaces [1]
2. Fisherfaces [2]
3. VGG Face net [3]

Representative of **classical approaches**.

Only CNN method with **official code** made **public**.

[1] - M. Turk and A. Pentland, “Eigenfaces for recognition,” J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71–86, Jan. 1991.
[2] - P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection.” pp. 711–720, 1997.
[3] - O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition.” in BMVC, X. Xie, M. W. Jones, and G. K. L. Tam, Eds. BMVA Press, 2015, pp. 41.1–41.12.

Eigenfaces

Eigenfaces seeks to **maximize data scatter**, and find the principal vectors to describe the image space.

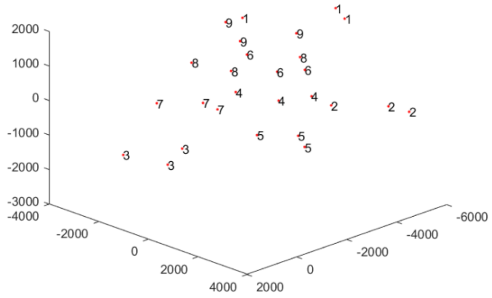
A new image **x** is **projected** into the face space by:

$$\omega_k = \mathbf{u}_k^T(x - \Psi) \qquad k = 1, \ldots, M'$$

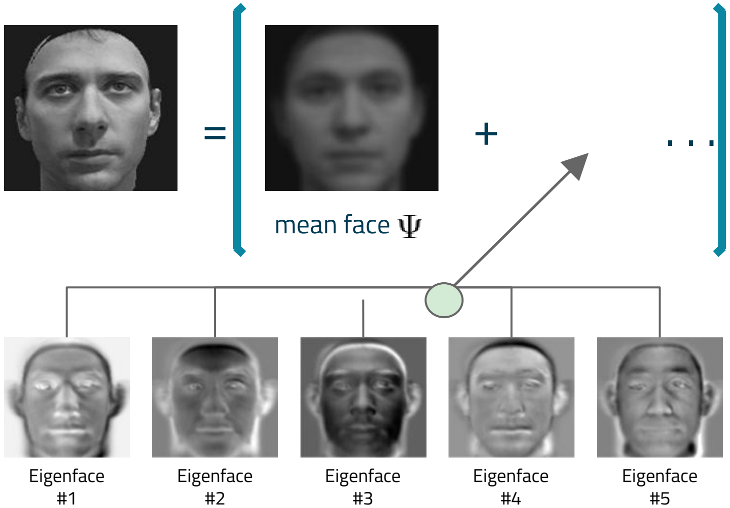
With $\mathbf{u}_k$ an eigenface.

This gives the **descriptor** by the weight vector

$$\Omega^T = [\omega_1, \omega_2, \ldots, \omega_{M'}]$$



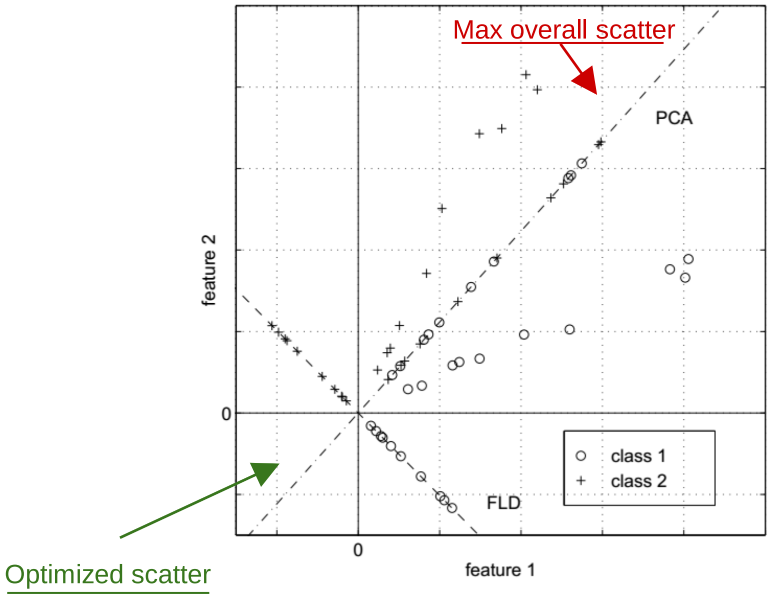
Face space



Fisherfaces

Fisherfaces adds the condition of **maximizing** the **ratio** of the determinant of the **between-class scatter** matrix of projected samples, **to** the determinant of the **within-class scatter** matrix of the projected samples.

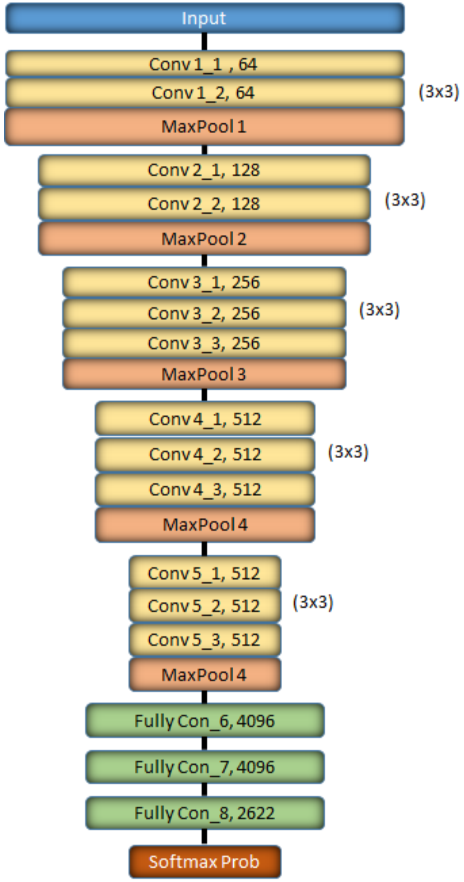
$$\Omega_{opt} = \arg\max_{\Omega} \frac{|\Omega^T S_B \Omega|}{|\Omega^T S_W \Omega|} = [\omega_1, \omega_2, \ldots, \omega_{M'}]$$



VGG Face net

1. Based on the CNN from **2014 Simonyan and Zisserman's** proposal.
2. **VGG Face** was trained specifically to **identify 2622 subjects** from a **custom built database** on celebrities' face pictures [1].
3. With **triplet loss training** reaches **0.9913** accuracy on **LFW** dataset.

Triplet training loss used in [1]



VGG Face architecture

[1] - O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition." in BMVC, X. Xie, M. W. Jones, and G. K. L. Tam, Eds. BMVA Press, 2015, pp. 41.1–41.12.

Experiments

Hardware platform:

- Intel **Core i7** 6500U CPU processor with **2.6 GHz** clock, **12 GB DDR3L** 1600 MHz SDRAM and one **NVIDIA GeForce GT 940M** 2GB DDR3.

Program language:

- **Eigenfaces** and **Fisherfaces** programmed on **Matlab R2016b** for Windows.
- **VGG Face** programmed on **Pycaffe** for Linux.

Dataset preparations:

All samples were processed with **frontalization**, cropping the **face detection** after.

Original 20°

Frontalized

Cropped detection

MIT-CBCL database used.

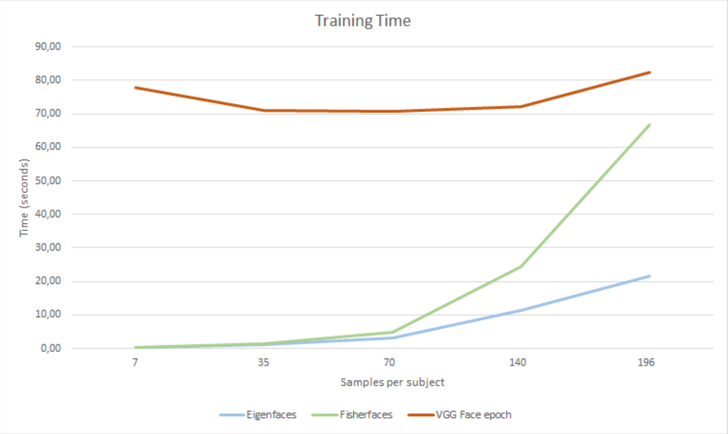
Results

Training Time

Samples per subject	Eigenfaces(s)	Fisherfaces(s)	VGG Face (min)
7	0.16	0.37	129.86 (77.92s)
35	1.09	1.46	177.28 (70.91s)
70	3.10	4.71	176,75 (70.7s)
140	11.26	24.40	180,24 (72,09s)
196	21.58	66.92	205,93 (82,37s)

Prediction Time

Eigenfaces	Fisherfaces	VGG Face
1.0765 s	1.0517 s	7.1067 s



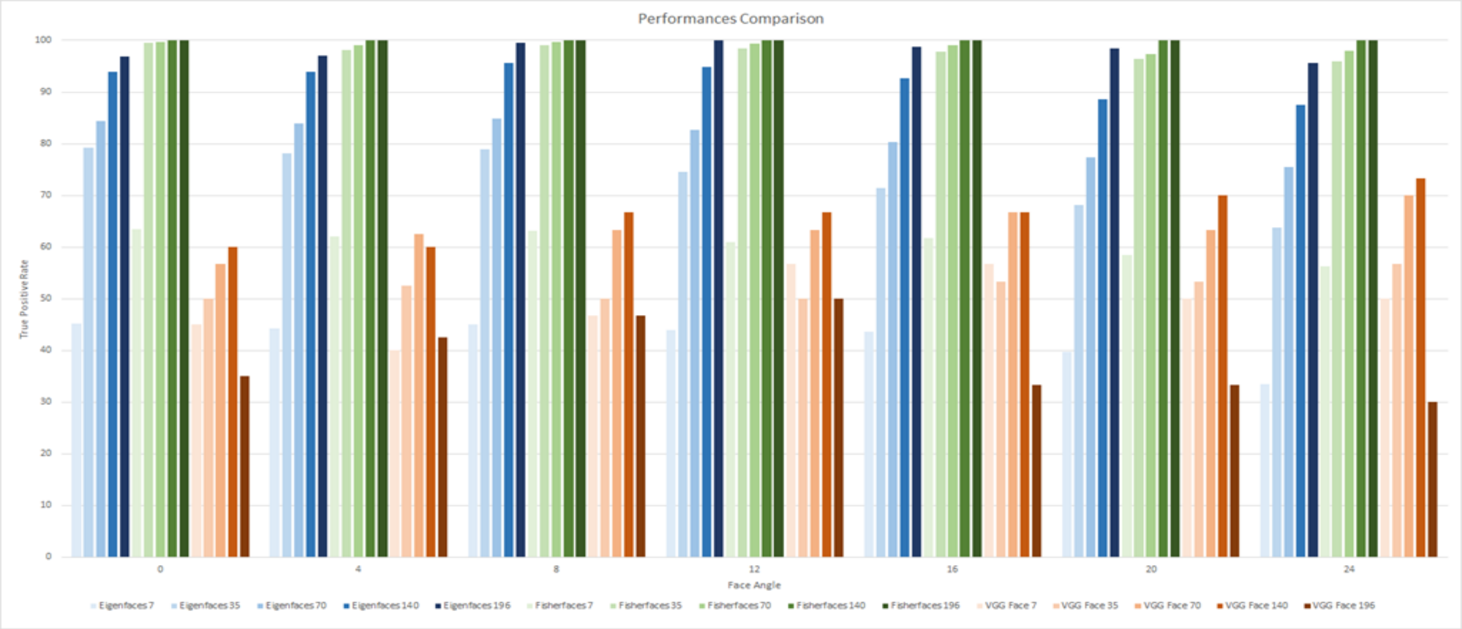
Results

Overall Performance

Samples per subject	Eigenfaces %	Fisherfaces %	VGG Face %
7	42.14	60.85	48.4
35	73.43	97.9	52
70	81.28	98.87	**62.8**
140	92.42	**100**	65.2
196	**98.04**	100	38.4

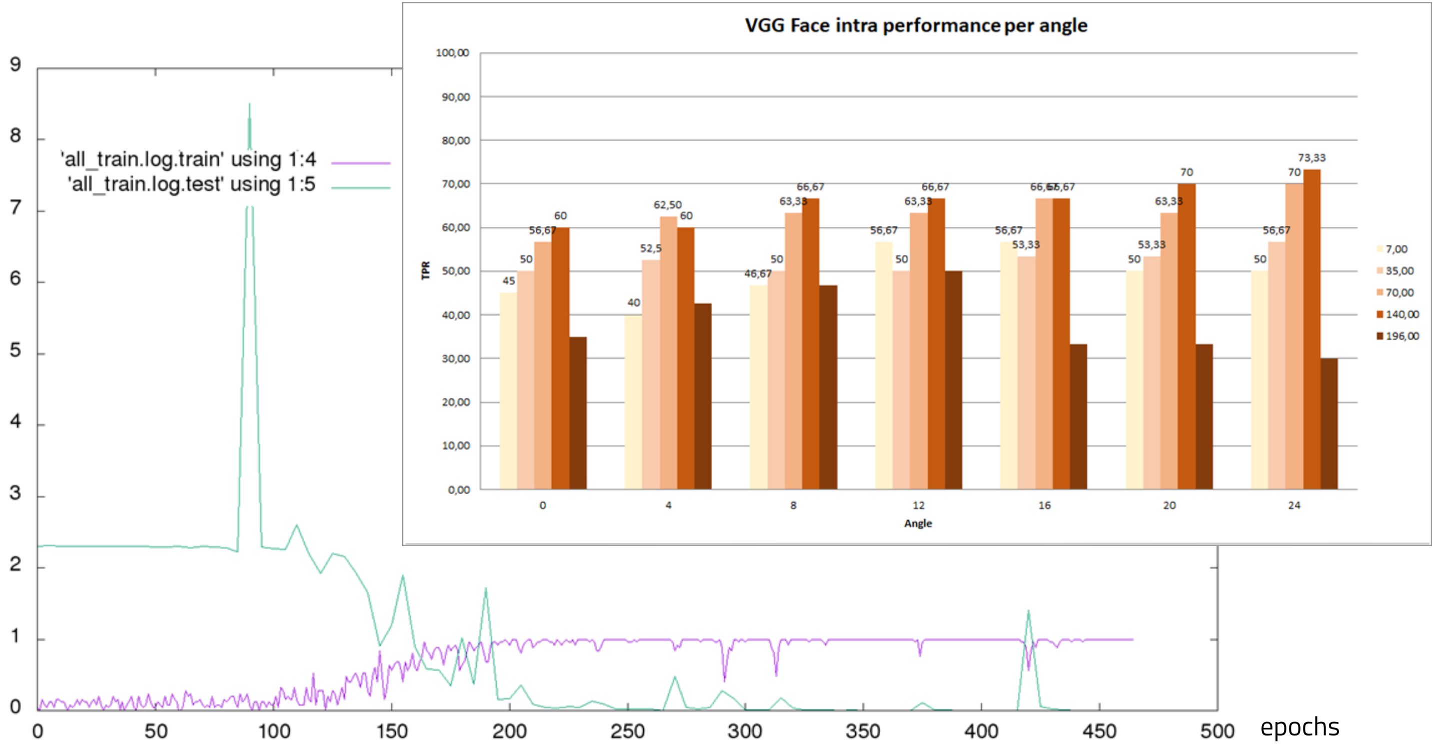
Face recognition with **train and test sets splits** over a database of **frontalized faces** using a **10-folds scheme**.

Results



Performance comparison for each face rotation angle between the 3 models, when varying samples per subject. Eigenfaces in blues, Fisherfaces in greens, VGG Face in browns.

VGG overfitting observations



Conclusions and future work

- Reached **near real-time** recognition with **over 95%** accuracy (**Eigenfaces and Fisherfaces**) for non optimized code, on **MIT-CBCL database**.
- **Viable field tests** with chosen methods.
- MIT-CBCL **dataset** is too **small** for **VGG Face** and facilitates **overfitting** without intense **supervision**.

Future work includes:

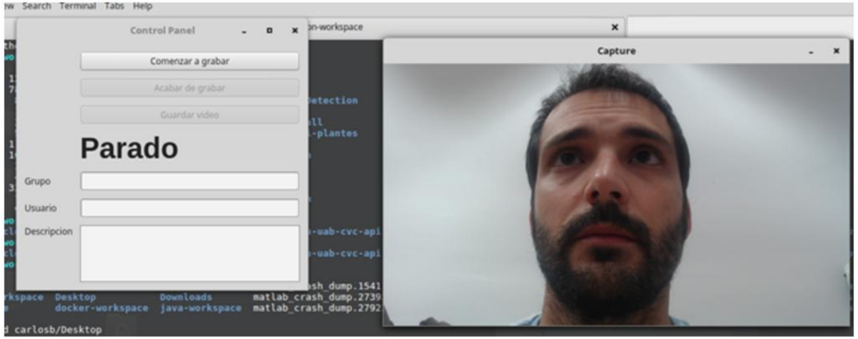
- Construction of **FCB's athletes face database**.
- VGG Face **triplet loss training** to boost **performance**.
- **Tests** on data with **more subjects**.
- **Tests** on **simultaneous recognitions** to define final **hardware** requirements.

Appendix

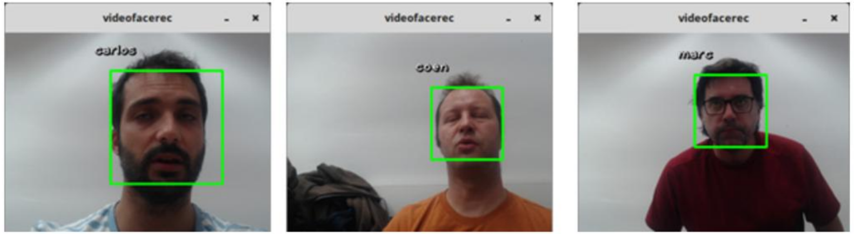
Experimental setup for face recognition in **FCB's gym** facilities based on **Linux and Raspberry Pi** platform was developed.

It includes the cameras, server and dedicated application **to build FCB's database**.

Raspberry Pi unit



Application to build FCB's gym database.



Tests for face recognition with the setup in laboratory.

Thank you!

Aknowledgements:

To **F.C. Barcelona** for their support and disposition to carry this collaboration project, in particular to **Raul Peláez Blanco** and **Joan Ramón Tarragó**.

To **Carlos Báez** and **Coen Antens** from **CVC**.

Questions?