Recognition of Occluded and Lateral Faces Using MTCNN, DLIB and Homographies

Gustavo A. Bezerra (Koruja)

November 26, 2018

Programa de Educação Tutorial - Ciência da Computação
Departamento de Matemática e Informática Aplicada
Universidade Federal do Rio Grande do Norte

Table of contents

Introduction

- SIBGRAPI 2018;
- Detailed work;
- DIM0097 - Tópicos Especiais em Computação VIII;
- More to come?

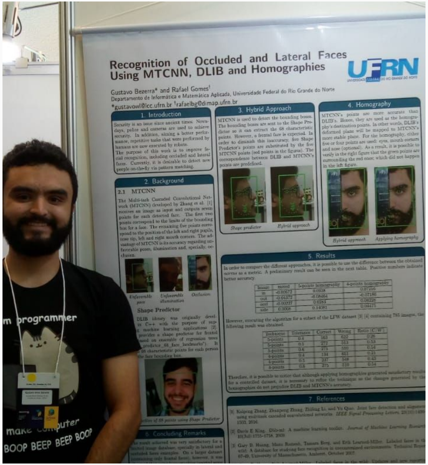


Figure 1: Poster presentation at SIBGRAPI 2018

Motivation

Security - Outside → Inside

(a) Great Wall of China

(b) Guards/Police

(c) Cameras

Figure 2: Security Measures

Motivation - Problem

People monitoring/tracking

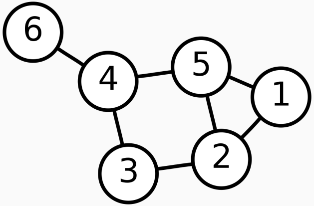


Figure 3: Undirected Graph

Figure 4: Instituto Metrópole Digital

Issues - Too many people making too many problems

- Tracking;
- Object recognition;
- People recognition;
- Face recognition.

- Small groups;
- Time limit;
- Few resourcers;
- Undergraduate course.

Solution - Improvise. Adapt. Overcome.

- Reduce problem's scope;
- Deep Neural Networks;
- MTCNN & DLIB;
- Comparison between initial and final results;

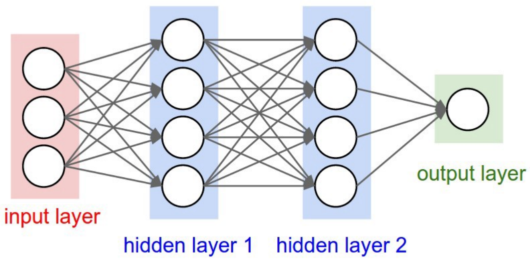


Figure 5: An example of a "Deep" Neural Network

Issues - Too many people making too many problems (again)

- Satisfactory initial results;
- Core problem to solve: reduce scope one more time.

Related Works

Dense 3D Alignment from 2D Videos in Real time

- Range of approximately 60 degrees;
- Dataset of 3D faces;
- 3D meshes;
- Robustness for illumination;
- 2D to 3D annotations;
- Reconstruct 3D meshes.

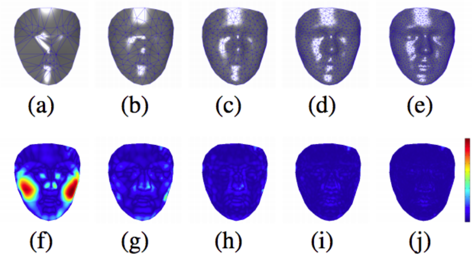
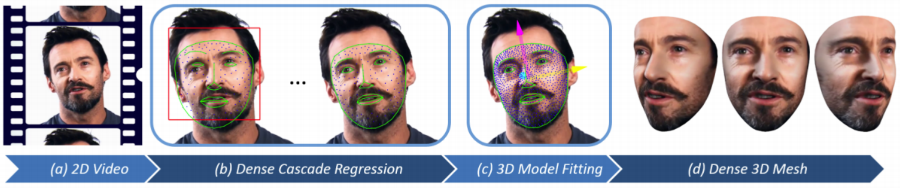
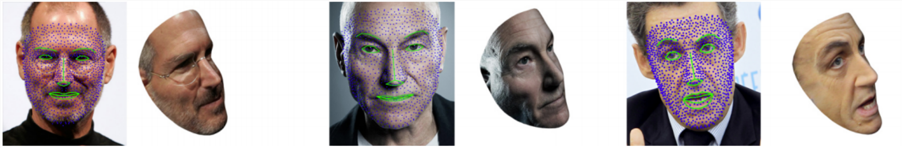


Figure 6: Reconstruction of 3D meshes

Dense 3D Alignment from 2D Videos in Real time



(a) Algorithm outline



(b) Meshes generated for celebrities

Figure 7: More examples of Dense 3D meshes generated

Continuous Supervised Descent Method for Facial Landmark Localisation

- Small and fast training models;
- Synthesised largest 2D dataset;
    - Challenging rotations;
    - 3D $\rightarrow$ 2D;
- Synthesising process;
    - Generate 3D face model;
    - Rotate face;
    - compute projection;
    - Add it to dataset (with texture).

Figure 8: Dataset synthesised images example

Continuous Supervised Descent Method for Facial Landmark Localisation

Lacks format consistency.

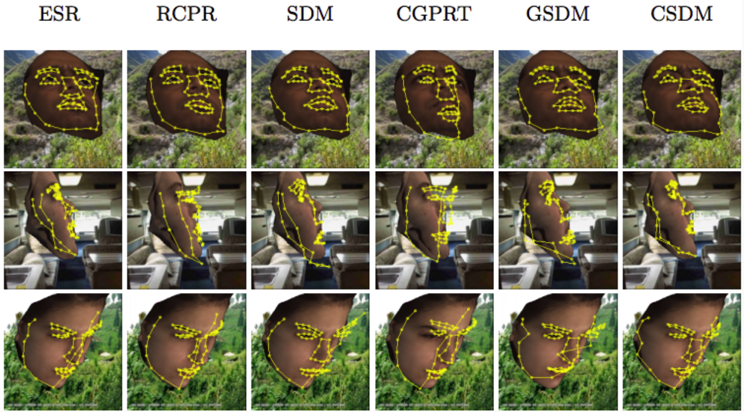


Figure 9: CSDM Comparison with other techniques

Viewpoint-Consistent 3D Face Alignment

- Consistent 3D view from 2D images;
- New dataset;
- New paradigm;
  - New Challenge;
  - No previous work validating consistency;
- Full 3D;
  - Outputs 3D points, not 2D.

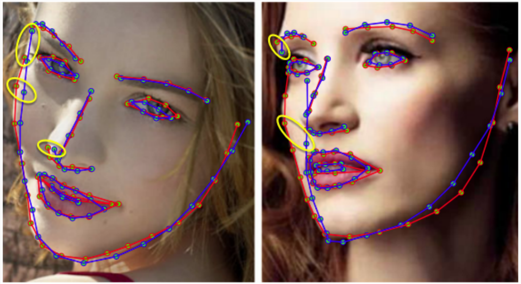


Figure 10: Comparison between 2D inconsistent and 3D consistent points

Viewpoint-Consistent 3D Face Alignment

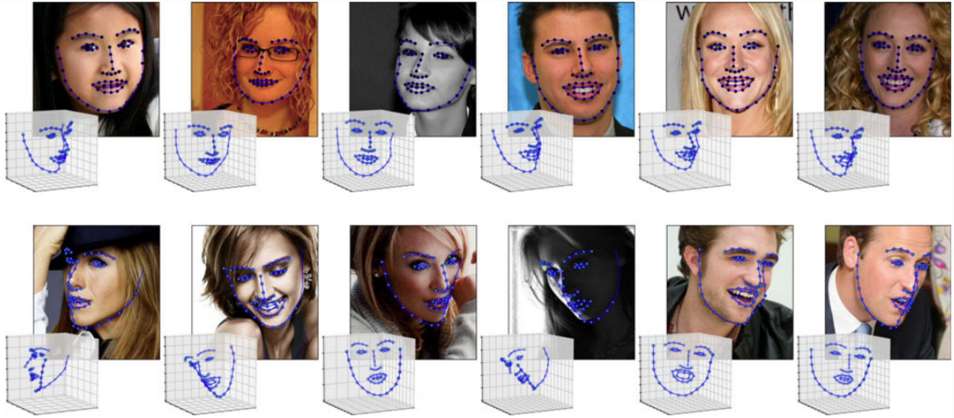


Figure 11: Some results obtained

Proposed Solution

Proposed Solution

- Tackle extreme conditions (with available tools);
  - Unfavourable illumination;
  - Unfavourable poses;
  - Object occlusion;
- Pre-trained DNN;
  - Shape Predictor (DLIB);
  - MTCNN;
  - ResNet.

Shape Predictor (DLIB)

- shape_predictor_68_face_landmarks;
- Ensemble of Regression Trees;
- Fit a generic face model.

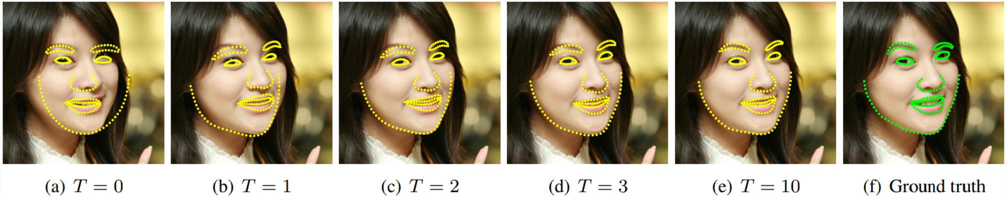
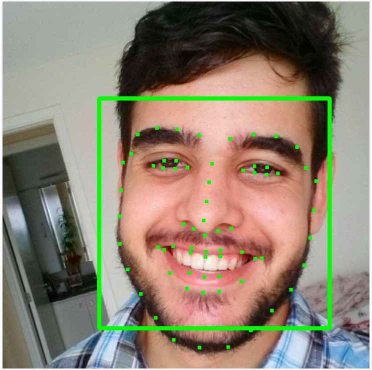


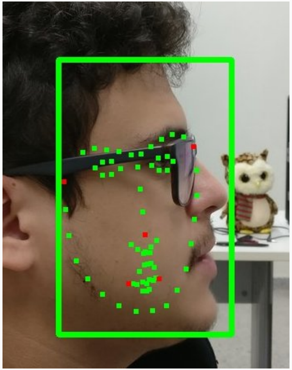
Figure 12: Ensemble of Regression Trees in action

Issues - Too many people making too many problems (one more time)

Shape Predictor works properly for frontal faces only.



(a) Frontal face



(b) Lateral face

Figure 13: Shape Predictor attempting to fit faces in different poses

MTCNN

- Stands for Multi-task Cascaded Convolutional Network;
- Uses three CNNs with different purposes;
  - Proposal;
  - Refine;
  - Final bounding box and landmarks.

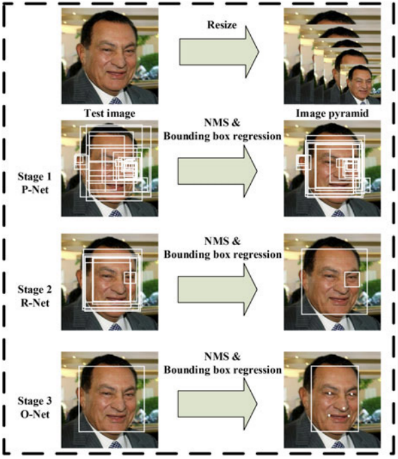
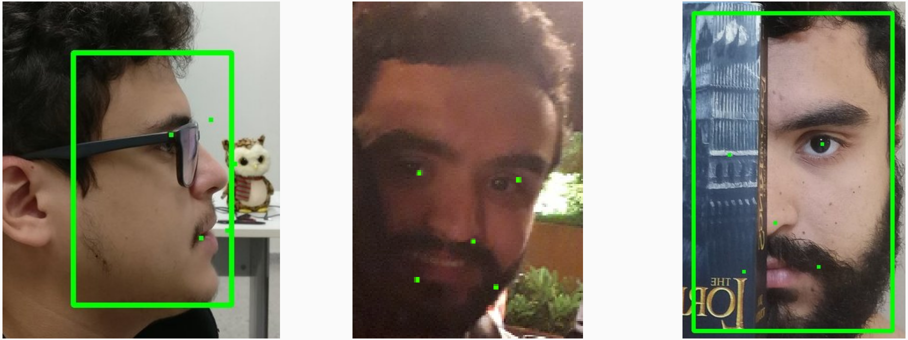


Figure 14: MTCNN outline

MTCNN works well for extreme conditions.



(a) Unfavourable pose **(b)** Unfavourable illumination **(c)** Occluded face

Figure 15: Examples of MTCNN detecting faces in extreme conditions

Issues - Too many people making too many problems (once more)

More points are necessary to recognise a person's face.

ResNet

- $f : \mathbb{R}^{68^2} \to \mathbb{R}^{128}$;
- Characteristic vector (Unitary and $\mathbf{v} \in \mathbb{R}^{128}$);
- Vector used for comparison;
- Similarity depends on *tolerance* factor;
- Accuracy of 99.38**%** claimed for Labeled Faces in The Wild dataset.

ResNet - Architecture

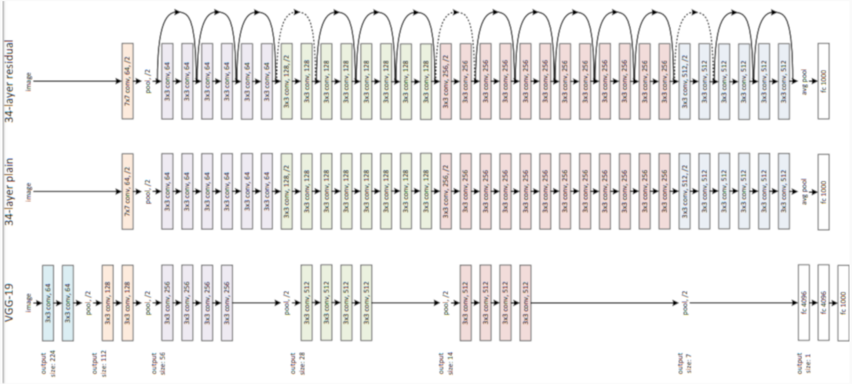
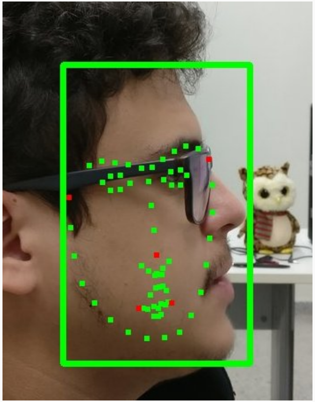
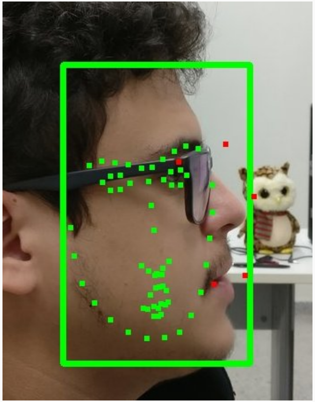


Figure 16: Architecture comparison between VGG-19 (bottom), 34-layer CNN (middle), and 34-layer ResNet (top)

Hybrid Approach - MTCNN and Shape Predictor



(a) Shape Predictor only



(b) Hybrid approach

Figure 17: Five Shape Predictor's points are substituted by MTCNN's

Issues - Too many people making too many problems (count++)

- Five MTCNN points were not sufficient;
- Five MTCNN points won't be sufficient.

Homography - Concept

- Simulate view from another position;
- Map position of four target and destiny points;
- Convert plane A to plane B.

Homography - Example

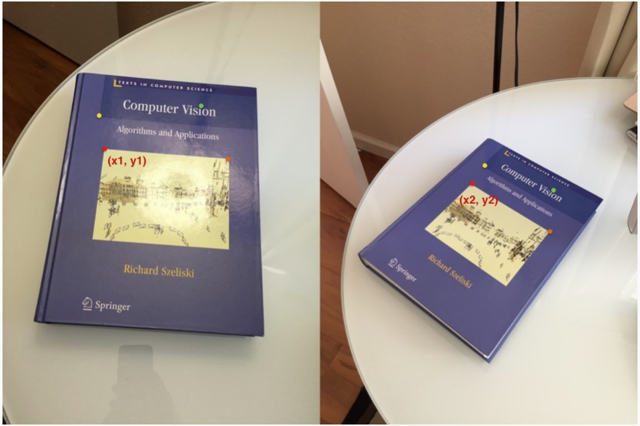


Figure 18: Two pictures of the same object from different points

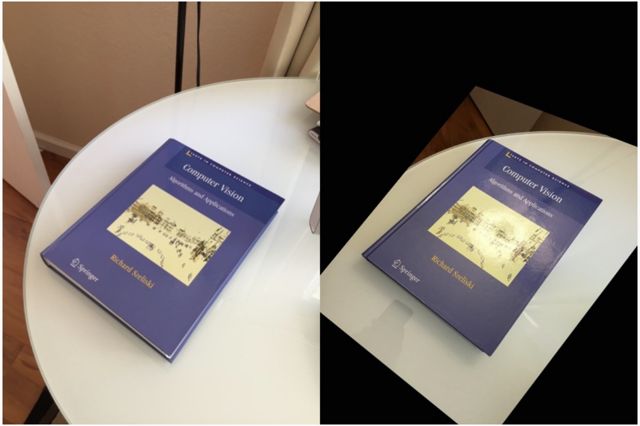


Figure 19: Comparison between a photo and an image generated via homography

- Use Shape Predictor points as targets;
- Use MTCNN points as destiny;
  - Nose is optional;
- Apply homography.

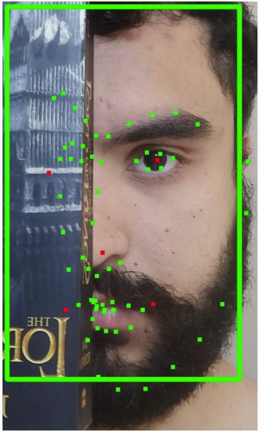


Figure 20: Occluded face using hybrid approach

Homography - Application

- Use Shape Predictor points as targets;
- Use MTCNN points as destiny;
  - Nose is optional;
- Apply homography.

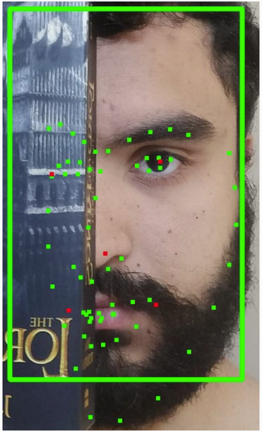


Figure 21: Occluded face using homography

Results

Results

Image	mixed	5-points homography	4-points homography
in	-0.00672	0.0938	0.07216
out	-0.04372	-0.08464	-0.07186
occl	-0.00227	0.0283	0.08228
side	0.0068	0.14081	0.04475

Table 1: Results for a small set of unfavourable images

Results

Technique	Tolerance	Correct	Wrong	Ratio (C/W)
5-points	0.4	163	622	0.26
5-points	0.5	272	513	0.53
5-points	0.6	275	510	0.54
4-points	0.4	134	651	0.21
4-points	0.5	237	548	0.43
4-points	0.6	275	510	0.54

Table 2: Results for Labeled Faces in the Wild dataset

Conclusion

Conclusion

- Results satisfactory for a limited dataset;
- Noise insertion due to chin elongation;
- It's a long way to the top if you wanna rock 'n' roll.

K. He, X. Zhang, S. Ren, and J. Sun.
Deep residual learning for image recognition.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

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.

L. A. Jeni, J. F. Cohn, and T. Kanade.
Dense 3d face alignment from 2d videos in real-time.
In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, volume 1, pages 1–8. IEEE, 2015.

D. E. King.
Dlib-ml: A machine learning toolkit.
Journal of Machine Learning Research, 10(Jul):1755–1758, 2009.

M. Koestinger, P. Wohlhart, P. M. Roth, and H. Bischof.
Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization.
In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 2144–2151. IEEE, 2011.

G. B. H. E. Learned-Miller.
Labeled faces in the wild: Updates and new reporting procedures.
Technical Report UM-CS-2014-003, University of Massachusetts, Amherst, May 2014.

M. Oliu, C. Corneanu, L. A. Jeni, J. F. Cohn, T. Kanade, and S. Escalera.
Continuous supervised descent method for facial landmark localisation.
In *Asian Conference on Computer Vision*, pages 121–135. Springer, 2016.

J. Shen, S. Zafeiriou, G. Chrysos, J. Kossaifi, G. Tzimiropoulos, and M. Pantic.
The first facial landmark tracking in-the-wild challenge: Benchmark and results.
pages 1003–1011, 2015.

S. Tulyakov, L. A. Jeni, J. F. Cohn, and N. Sebe.
consistent 3d face alignment.
IEEE transactions on pattern analysis and machine intelligence, 40(9):2250–2264, 2018.

S. Yang, P. Luo, C. Loy, and X. Tang.
Wider face: A face detection benchmark.
pages 5525–5533, 2016.

- S. Zafeiriou, G. Trigeorgis, G. Chrysos, J. Deng, and J. Shen.
  The menpo facial landmark localisation challenge: A step towards the solution.
  pages 2116–2125, 2017.
- K. Zhang, Z. Zhang, Z. Li, and Y. Qiao.
  Joint face detection and alignment using multitask cascaded convolutional networks.
  IEEE Signal Processing Letters, 23(10):1499–1503, 2016.
- Z. Zhang, P. Luo, C. C. Loy, and X. Tang.
  Learning deep representation for face alignment with auxiliary attributes.
  IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(5):918–930, May 2016.

Questions?

Questions?