

# **Facial detection and landmarking**

**University of York, Department of Computer Science**

**Yuri Pieters**

**8th June 2022**

# **Contents**

<b>Executive Summary</b>	<b>4</b>
<b>Introduction</b>	<b>5</b>
Notes . . . . .	5
Introduction . . . . .	6
<b>Methods</b>	<b>10</b>
<b>Results</b>	<b>11</b>
<b>Conclusion</b>	<b>12</b>
<b>Bibliography</b>	<b>13</b>

# List of Figures

1	Example of a face from the 300W face database [2] with a set of 68 landmark points annotated. The 68 point pattern was first used for the Multi-PIE [3] database, but has since been used on many other databases [2]. . . . .	7
2	Two images from the 300w dataset, showing some of the range of human faces. . . . .	9

# Executive Summary

- State aim
- Motivation: why is worth doing this?
- Methods: how was it done?
- Results: what was found
- Issues: what should people be mindful of with regards to this subject?

The goal of this work was to understand, compare, and evaluate a range of different techniques used in the detection and landmarking of human faces by computers. In this report I present the models I chose to evaluate, the datasets I evaluated them against, and the results of the evaluation. I also present a review of the field overall, which motivates the particular models and datasets I choose.

Motivation:

- Computer vision is an important part of many topics at the bleeding edge of computer science: robotics, self driving cars, human computer interfaces.
- Computer vision techniques aren't one size fits all. To pick between them requires understanding of the differences.
- Recognition of faces is an important and frequently used application of computer vision.
- Therefore understanding and picking between face recognition algorithms is important.

Social and ethical issues of face detection and landmarking:

- Human faces have a huge variety
- Everyone has a right to be able to use the technology that may be supported by computer vision
- Therefore both the computer vision techniques and actual implementation of those techniques (e.g. the actual training images used) should be developed with the full diversity of humans in mind.

# Introduction

## Notes

- Subject area intro/background
- Narrow to specific background for this work
- Restate motivation
- Restate aims
  
- Object detection and landmarking are general problems. Faces are a popular subject though.
- Overall aim of face detection/landmarking algorithms
- Difference between statistical models and neural networks
- Stages of a typical algorithm
- Stages in detail (I think it's detection then landmarking)
  - Detection
    - \* Overview of model categories
    - \* What is state of the art like?
  - Landmarking
    - \* Overview of model categories
    - \* What is state of the art like?
  
- What particular problems often occur with face detection/landmarking?
  - Occlusion
  - Extreme poses
  - More varied faces than the detector was trained on

Aims:

- Investigate a variety of techniques used in face detection and landmarking
- Evaluate a cross-section of these techniques using a unified data set

Face analysis categories:

- Face detection: answer's the question "is there a face in this image". Can typically be extended to answer "where are the faces in this image" by running the detector on all sub-regions of the image.
- Facial landmarking: going beyond detecting the presence of a face, landmarking identifies the position of the individual face parts. Different techniques use different numbers of landmarks. Some aim to fit a contour around every part, and some may just place points on the eyes, nose and mouth. The landmarks can be used as a precursor to other face processing tasks such as recognising emotions or identifying individuals.

## Introduction

The problems of face detection and landmarking are related but typically distinct problems in the field of computer vision. The goal of face detection is to detect the presence of faces in an image, often with the further goal of placing a bounding box around the face; this is a good start for other face analysis algorithms which assume the presence of a face, and is used in applications such as cameras to help auto-focus. The goal of landmarking a face is to place a set of points onto key parts of the face, as in fig. 1; the landmarks could then be used in further processing, such as identity recognition, facial behaviour analysis, lip reading, 3D face reconstruction, or face editing, to name a few examples [1].

Historically the problems of face detection and landmarking were tackled separately; however more recent work, such as that done by Zhu and Ramanan [4], has resulted in algorithms unifying both goals. In addition, the performance of some algorithms for landmarking is strongly affected by where they are initialised, and this initialisation is done with a face detector. To get the best performance in this case, the detection algorithm should be tailored to the one for landmarking [2].

In this project different off-the-shelf algorithms for facial detection and landmarking are compared, examining their performance at different image resolutions.

The 300 Faces In-The-Wild challenge [2] and similar competitions ([1], [5], [6]) have compared a variety of



Figure 1: Example of a face from the 300W face database [2] with a set of 68 landmark points annotated. The 68 point pattern was first used for the Multi-PIE [3] database, but has since been used on many other databases [2].

While they are special cases of the more general problems in object recognition, faces are frequently the case chosen in the literature, even when the algorithm in question is more generally applicable (e.g [7]–[9]). This may just be because it is easier to compare new work to old when using the same class of objects, but faces are also a generally good class to use for this comparison: they are non-rigid and diverse, making them a relatively difficult problem, and it is easy for a human to qualitatively judge the result.

Faces hold a special place in the human mind. As highly visual, social creatures our faces are an important part of how we navigate the world. With our faces we can communicate complex emotion, and they are one of the primary ways we recognise one another. Therefore the ability to recognise faces is hard-wired into our brains, so much so that we'll happily see them in everything from burnt toast to clouds. Its no wonder then that faces are a popular subject for computer vision research. However, writing a computer program to analyse human faces (or images of objects in general, actually) turns out to be easier said than done, and so has been the subject of much research over the last few decades. This research has produced many practical techniques already in use around us, such as face detectors helping to auto-focus your camera, or auto-tagging of identities in photo libraries. There is still much room for improvement, as most of these programs are still easily outperformed by humans. Research continues, however, and the state of the art continues to get closer to human level every year.

The goal of facial recognition is a broad one, and can be broken down further. Some tasks that can be included in facial recognition are

**Face detection** Detect the presence of a face in in image. The result may be a bounding box around the face, and is a good starting point for further analysis which assumes the presence of a face.

**Classification** Classify faces into categories such as gender, age, skin color, etc.



Figure 2: Two images from the 300w dataset, showing some of the range of human faces.

# **Methods**

- Approach
- What was done
- Justification for above; include reference to ethics.

# Results

- What happened
- Analysis of what happened

# **Conclusion**

- Final wrap up what happened
- Project aims

# Bibliography

- [1] J. Deng *et al.*, “The Menpo Benchmark for Multi-pose 2D and 3D Facial Landmark Localisation and Tracking,” *Int J Comput Vis*, vol. 127, no. 6, pp. 599–624, Jun. 2019, doi: [10.1007/s11263-018-1134-y](https://doi.org/10.1007/s11263-018-1134-y).
- [2] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, “300 Faces In-The-Wild Challenge: database and results,” *Image and Vision Computing*, vol. 47, pp. 3–18, Mar. 2016, doi: [10.1016/j.imavis.2016.01.002](https://doi.org/10.1016/j.imavis.2016.01.002).
- [3] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-PIE,” *Proc Int Conf Autom Face Gesture Recognit*, vol. 28, no. 5, pp. 807–813, May 2010, doi: [10.1016/j.imavis.2009.08.002](https://doi.org/10.1016/j.imavis.2009.08.002).
- [4] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 2879–2886. doi: [10.1109/CVPR.2012.6248014](https://doi.org/10.1109/CVPR.2012.6248014).
- [5] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The Pascal Visual Object Classes Challenge: A Retrospective,” *Int J Comput Vis*, vol. 111, no. 1, pp. 98–136, Jan. 2015, doi: [10.1007/s11263-014-0733-5](https://doi.org/10.1007/s11263-014-0733-5).
- [6] H. Nada, V. A. Sindagi, H. Zhang, and V. M. Patel, “Pushing the Limits of Unconstrained Face Detection: a Challenge Dataset and Baseline Results,” in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Oct. 2018, pp. 1–10. doi: [10.1109/BTAS.2018.8698561](https://doi.org/10.1109/BTAS.2018.8698561).
- [7] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, “Robust Discriminative Response Map Fitting with Constrained Local Models,” 2013, pp. 3444–3451. Accessed: Nov. 23, 2021. [Online]. Available: [https://openaccess.thecvf.com/content\\_cvpr\\_2013/html/Asthana\\_Robust\\_Discriminative\\_Response\\_2013\\_CVP\\_R\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2013/html/Asthana_Robust_Discriminative_Response_2013_CVP_R_paper.html)
- [8] J. M. Saragih, S. Lucey, and J. F. Cohn, “Deformable Model Fitting by Regularized Landmark Mean-Shift,” *Int J Comput Vis*, vol. 91, no. 2, pp. 200–215, Jan. 2011, doi: [10.1007/s11263-010-0380-4](https://doi.org/10.1007/s11263-010-0380-4).

- [9] S. Ioffe and D. Forsyth, "Mixtures of trees for object recognition," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, Dec. 2001, vol. 2, pp. II-180-II-185. doi: [10.1109/CVPR.2001.990953](https://doi.org/10.1109/CVPR.2001.990953).