Department of Computer Science

Submitted in part fulfilment for the degree of BSc.

Evaulation of facial landmarking methods

Yuri Pieters

DRAFT - May 11, 2022

Supervisor: Nick Pears

Contents

List of Figures

Executive Summary

the following is an outline

Aim: The goal of this work was to understand, compare, and evaluate a range of different techniques used for landmark localisation on faces. 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.

Methods: I evaluated the algorithms by running them on the same set of data, and calculating how closely they reproduced the true landmark points.

Results: I found that ….

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.

1 Introduction

Automatic analysis of human faces is a complex problem. Humans have an intuitive understanding of the human face from a very young age (fig. 1.1), and we quickly learn to interpret the faces of others to tell us who they are, where they are looking, what they are feeling, and more; the face is a rich form of non-verbal communication [1]. The ease with which we do this belies the difficulty we have had in teaching computers to do the same, however. Trying to infer something as subtle as emotion from a collection of pixels, full of noise arising from lighting conditions, camera properties, accessories such as hats and glasses, or simply the diversity of human faces, is a non-trivial task. The solution, of course, is to pre-process the data somehow to extract only relevant features [2]. This breaks a difficult problem (such analysing facial expressions), into more manageable sub-problems, which can be tackled separately. One such pre-processing technique that turns out to be useful for multiple different tasks is facial landmarking [2], [3].

Figure 1.1: An expert face analyser[1]

The goal of facial landmarking is to fit a parameterised shape model to an image of a face such that its points correspond to consistent locations on the face [4]. Fig. 1.2 shows an example of a face fitted with such a model. Landmarks points may either correspond to well defined facial part, such as the tip of the nose or the corner of the eye, or they

[1] "Sleeping Baby" (CC BY-NC-SA 2.0) by bikesandwich on Flickr

may be part of a group marking a boundary, such as the edge of the face. The exact configuration and meaning of the landmark points can vary between datasets used, and several different configurations are in use. One of the most popular however is the 68 point annotation used originally by the Multi-PIE dataset [5].

Figure 1.2: Example of a face from the 300W face dataset [6] with a set of landmark points annotated.

Facial landmarking can be used to as part of the process of solving more difficult facial analysis problems in different ways. Most obviously, landmarks can be used directly as input data for a model. A model for head pose estimation for example may not actually need most of the information encoded in the image pixels; the pose of the head can be inferred from the relative positions of the various facial features, which is what the landmarks encode [3]. Alternatively, the landmarks can be used as part of additional pre-processing steps such as registration and feature extraction [2]. In registration the idea is to remove variation in rotation and scale; this can be done by first computing a transformation that places the landmarks onto a predefined reference shape, and then applying the same transformation to the image itself. In feature extraction the goal is to compute summaries of the image data that keeps relevant information while getting rid of nuisance factors. Landmarks can help localise the features, so that each feature represents the same part of the face in every example. Features can also be computed directly on the landmarks themselves, encoding geometric relationships between different parts of the face.

In the following, I present an overview of facial landmarking methods, followed by a more detailed explanation of a few specific algorithms. Those algorithms are then evaluated against a common dataset, before

concluding with with some recommendations on the suitability of the tested algorithms in certain circumstances.

1.1 Landmarking configurations

consider moving or rewriting this

Before going deeper into methods of computing landmarks, I shall talk briefly about landmark configurations. Different landmarking schemes exist, with varying numbers of landmarks and with the landmarks localised to different parts of the face [6]. The different configurations have arisen from the different decisions made by people collecting and annotating datasets, and appear to be mostly arbitrary. One of the most widely used is the 68 point configuration from the Multi-PIE dataset (which is the one shown in fig. 1.2). Thanks to the work done for the 300 Faces In-The-Wild challenge [6], there are now annotations using the 68 point configuration available for several public datasets.

1.2 Overview of landmarking methods

There have been many approaches taken to the problem of landmark fitting, but they can largely be divided into three categories [7]: *holistic methods*, *Constrained Local Models*, and *regression based methods*. The categories are based on how the facial appearance and facial shape patterns are modelled and related. In the holistic methods, a model of the global appearance of the face is related to a global model of the landmark positions. Constrained Local Model (CLM) approaches train a set of independent models for each of the facial landmarks, but constrain the locations of the landmarks based on a global model of the face shape. Lastly, the regression based methods do not explicitly model the global face shape at all, instead directly relating image data (either local or global) to landmark locations.

1.2.1 Holistic methods

The holistic methods for facial landmarking are a family of generative statistical models, whose prototypical version is known as Active Ap-

pearance Models (AAM), as proposed by Cootes et al [8]. AAM works by learning two connected models using Principle Component Analysis (PCA): a model of face shape, and a model of global face appearance. These are connected through sharing the same set of parameters; the joint parameterisation allows the model to capture appearance variation due to shape, such as teeth appearing when the mouth is open [9]. To simplify the model and make it more robust, Procrustes analysis [10] is used on the training data to remove variation due to global transformations, and to find the mean face shape; the images are then warped so that each landmark is moved to its mean location, so that the appearance model is independent of shape. The goal when fitting the model to new images is to find both the optimal model parameters and the correct global transform. In the classic version of this model, fitting is done by iteratively calculating the error between the image as generated by the current parameters, and calculating an update to the parameters based on the error [7].

Discuss extensions to the algorithms

1.2.2 Constrained Local Models

The central idea of CLM is to model, for each landmark, the likelihood that it should be placed on a certain part of the image, but to then constrain the final landmark locations to fit a model of the face shape as a whole.

CLM models can be traced back work by Cristinacce and Cootes [11].
Describe basic form of CLM

Discuss extensions to the algorithms

1.2.3 Regression based methods

Describe class of algorithm and discuss extensions

2 Evaluations

2.1 Evaluated algorithms

- Holistic: [12]
- CLM: [13]
- Regression based: [14].

Give more details on each algorithm

2.2 Testing data

Choose subset of data not used for training any of the evaluated models. Describe the datasets.

The training data was divided between extreme and frontal poses. Many algorithms struggle when the face is not pointing mostly towards the camera. However, many situations in which face landmarking can be employed mostly involve frontal faces anyway; think of analysing faces in an online video call, for example. Therefore, good performance on frontal faces could still be valuable when coupled with good computational performance.

The segmentation between frontal and extreme poses was done in a semi-automated manner using an off-the-shelf head pose estimation algorithm. The algorithm used was from the OpenFace 2.0 toolkit [15], which uses the algorithm in [13] to fit a set of three-dimensional facial landmarks, from which rotation (and translation) relative to the camera is inferred. The accuracy is limited in this case, as it relies in part on knowing camera properties. Accuracy is not necessary in this situation, however, because the all that is needed is an approximation of frontal or extreme pose; nonetheless, manual verification of the process was performed. The procedure then for building the two subsets was:

1. Run the head pose estimation tool on the data
2. Classify into *frontal* and *extreme*:
    - Frontal was defined as a rotation in all directions less than *what?*
    - The remaining faces were classified as extreme. Note that this includes faces where the head-pose estimator returned no value.
3. Go through the two sets; reclassify any errors.

2.3 Evaluation criteria

The algorithms were evaluated on the point-to-point error between the fitted shape the ground truth annotations, normalised by the distance between the outer corners of the eye, as used in [6]. Specifically, given a set of N fitted landmark points $\mathbf{s}^f = \{\mathbf{x}_i^f\}_{i=1}^N$ and a corresponding set of N ground truth landmark points $\mathbf{s}^g = \{\mathbf{x}_i^g\}_{i=1}^N$, the error is:

$$\text{Error} = \frac{1}{d_{outer} N} \sum_{i=1}^{N} \|\mathbf{x}_i^f - \mathbf{x}_i^g\| \tag{2.1}$$

3 Results

- What happened
- Analysis of what happened

4 Conclusion

- Final wrap up what happened
- Project aims

Bibliography

[1] N. Kanwisher and G. Yovel, “Face Perception,” in *Handbook of Neuroscience for the Behavioral Sciences*, John Wiley & Sons, Ltd, 2009. doi: 10.1002/9780470478509.neubb002043.

[2] B. Martinez and M. F. Valstar, “Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition,” in *Advances in Face Detection and Facial Image Analysis*, M. Kawulok, M. E. Celebi, and B. Smolka, Eds. Cham: Springer International Publishing, 2016, pp. 63–100. doi: 10.1007/978-3-319-25958-1_4.

[3] E. Murphy-Chutorian and M. M. Trivedi, “Head Pose Estimation in Computer Vision: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607–626, Apr. 2009, doi: 10.1109/TPAMI.2008.106.

[4] 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.

[5] 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.

[6] 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.

[7] Y. Wu and Q. Ji, “Facial Landmark Detection: A Literature Survey,” *Int J Comput Vis*, vol. 127, no. 2, pp. 115–142, Feb. 2019, doi: 10.1007/s11263-018-1097-z.

[8] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, Jun. 2001, doi: 10.1109/34.927467.

[9] S. J. D. Prince, *Computer Vision: Models, Learning, and Inference*, 1st edition. New York: Cambridge University Press, 2012.

[10] J. C. Gower, “Generalized procrustes analysis,” *Psychometrika*, vol. 40, no. 1, pp. 33–51, Mar. 1975, doi: 10.1007/BF02291478.

[11] D. Cristinacce and T. F. Cootes, “Feature Detection and Tracking with Constrained Local Models,” in *Procedings of the British Machine Vision Conference 2006*, Edinburgh, 2006, pp. 95.1–95.10. doi: 10.5244/C.20.95.

[12] G. Tzimiropoulos and M. Pantic, “Optimization Problems for Fast AAM Fitting in-the-Wild,” in *2013 IEEE International Conference on Computer Vision*, Sydney, Australia, Dec. 2013, pp. 593–600. doi: 10.1109/ICCV.2013.79.

[13] A. Zadeh, T. Baltrusaitis, and L.-P. Morency, “Convolutional Experts Constrained Local Model for Facial Landmark Detection,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 2051–2059. doi: 10.1109/CVPRW.2017.256.

[14] V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 1867–1874. doi: 10.1109/CVPR.2014.241.

[15] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, “OpenFace 2.0: Facial Behavior Analysis Toolkit,” in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi’an, May 2018, pp. 59–66. doi: 10.1109/FG.2018.00019.