Student ID: 10358736
COMP61342

Detecting and Recognizing Faces in Images: A discussion on six contributions from 1991 to 2002

COMP61342 - The University of Manchester
Spring 2019

Being able to detect and recognise faces in images has have many valuable applications in today's world where increasing amounts of image data is shared and available. Applications of face detection and recognition are many, including human computer interaction, identification, and generally in image and film processing. The challenges which have been attempted solved through research for the last decades are several; including the initial face detection; whether there are faces in an image and where, and furthermore recognition; whether the face is known and to whom it belongs. Moreover, finding an algorithm which can overcome the challenges of variations in lightning, facial expression, pose, scale and limited computational resources has been pursued. This essay will discuss and compare six papers' contributions to advancements within the field of face detection and recognition.

The first of the contributions was *Eigenfaces for Recognition*, published in 1991 by Turk and Pentland. The paper defined the term *Eigenfaces*, and built upon the work of Sirovich and Kirby from 1987 whom had been pioneers in terms of how they applied linear algebra to the problem of facial recognition. Turk and Pentland defined *Eigenfaces* to be the set of eigenvectors (principal components) computed across a set of known face images. By capturing most of the variance among the faces this set of Eigenfaces then represented the key features in the different faces, allowing for face recognition task to be reduced to a problem in a lower dimension. A face could then be recognized by characterizing it as a weighted sum of the chosen Eigenface features. Overall, to recognize a certain face one only had to compare the weights to those face one already knew. Overall, their method further advanced the method of Sirovich and Kirby (1987) by also allowing to *detect* faces in images, and was an early example of a real-time facial detection and recognition method. They conducted experiments to assess the methods performance with slight variations in lightning, head size and head orientation and report to have achieved 96% correct classifications for cases with variations in lightning, 85% correct for variations in orientation and 64% correct with variations in size.

In 1997, Belhumer et al. published a paper which compared Turk and Pentland's Eigenfaces to a new method coined as *Fisherfaces*. Contrary to Eigenfaces, their methods make use of Fischer's Linear Discriminant analysis. Just as for Eigenface, the Fisherface method projected face images to a space of lower dimensions compared to the original images. However, the Eigenface method, using PCA, tried to maximize total scatter across classes of all images of all faces which resulted in keeping variations caused by lightning and facial expression. On the contrary, the Fischerfaces method selected projection directions to be close to orthogonal to the within face space scatter. As such, they try to maximize the ratio between the between-class variation and the within-class variation. Thus, they were able to rule out variations in lightning and facial expression, but still able to distinguish faces from each other. This resulted in the method obtaining better results under conditions of large variations in lightning and facial expressions. The method was tested on two databases of face images, while comparing results to other methods including correlation and Eigenfaces. The result show that the Fisherface methods performed equally well or better than the other models tested and especially under conditions of variations in lightning and facial expression.

The same year as the Fisherface method was introduced, Moghaddam and Pentland published their paper which combined Eigenfaces with a Probabilistic interpretation of the face detection and recognition challenge. Their model worked by sliding a probability “calculator” across images, calculating the probability that each sub image contains a face and then choosing the position, if any, in which it is the most likely that a face was present. To calculate the probability in each sub image they made use of PCA. However, contrary to many other applications using PCA, they included *both* the Principal Components, i.e. the eigenvectors which captures a majority of the variance in between faces, *and* the eigenvectors which are typically discarded - in the probability expression. They argue that this enables them to firstly detect if an image contains a face and secondly to whom the face belongs. As an example of performance, they find that their method, when tested on the FERET (NIST, 2019) database, is able to detect 97% of the faces correctly, overcoming challenges such as scale, position and occluding objects such as sunglasses

The year after, in 1998, the paper *Nonlinear Component Analysis as a Kernel Eigenvalue Problem* was published by Schölkopf et al. Where all the previous papers had made use of linear decomposition methods, Schölkopf et al. (1998) presented a further generalization of linear PCA to enable handling nonlinear problems. Their proposed method was a kernel based algorithm which mapped from input space to feature space using nonlinear mapping, and thereafter applied linear PCA to extract the principal components. Overall, this allowed them to capture higher order statistical dependencies in the image data. The experiments presented in the paper find that Kernel PCA obtained better recognition rates when using the same number of components as linear PCA and that Kernel PCA allowed for using more components than what could be accomplished with linear PCA. Moreover, compared to other nonlinear methods for feature extraction Schölkopf et al. (1998) argued that Kernel PCA had the advantage of just being a linear algebra problem to find eigenvalues, as opposed to other methods such as Neural Networks and Principal Curves which involved solving nonlinear optimization problems - which was much more complicated. Although Schölkopf et al. (1998) did not test their method on facial detection and recognition problems, later papers have showed that the Kernel PCA technique could outperform other popular methods at the time (Kim et al., 2002, Yang, 2002). For instance, Yang et al. (2002) find that Kernel Eigenfaces and Kernel Fisherfaces obtain lower error rates on facial recognition tasks than the traditional Eigenfaces and Fischerfaces approach.

In 2001, Viola and Jones presented a new method which stood out from the other papers both in terms of complexity, although the method is relatively easy to comprehend on an abstract level, and in terms of the speed in which it was able to detect faces. Firstly, Viola and Jones

(2001) introduce the integral image, a new image representation which allows for evaluating features very fast compared to previous methods. Secondly, they made use of a version of Adaboost to select the most important features in the images to be able to detect and recognize faces. Lastly, they present a model where they use several layers of classifiers with increasing complexity to detect interesting regions, trying to prevent spending too much time on less promising areas of images. When tested on face detection problems they report obtaining less than 1% false negatives and 40% false positives. The results from their test show that they are able to get high accuracy while operating 15 times faster than other detectors at the time, while handling typical challenges such as variations in lightning, scale and pose.

Last out of the papers is the contribution made by Vasilescu and Terzopoulos in 2002 which put forward the concept of *Tensorfaces*. The authors argued that linear analysis methods, such as methods using PCA and LDA, were not the best suited for recognising faces in the case of variations in lightning, pose, scale etc. Using a multilinear approach, i.e. using Tensors - the multidimensional generalization of a matrices, the authors represent the images of faces by higher order tensors than what had traditionally been done. By extending the conventional matrix singular value decomposition (SVD) to a higher order decompositions, they were able to handle problems related to lightning, pose, scale etc. which causes the image data to have a multimodal structure. Their experiments that their method outperforms the traditional linear methods like conventionat SVD and PCA.

As these papers where published across a span of ten year, it is natural that they tackle different challenges, and that the later ones profited from the advances and insights gathered by those before them. Keeping this in mind, a discussion of strengths and weaknesses of the approaches is interesting. Being the first of the methods put forward, the Eigenface method by Turk & Pentland (1991) can be argued to be simple and accurate relative to methods before it. It showed how PCA could be used to simplify the facial detection *and* recognition challenge in real-time. However, under small variations in scale and pose the performance was reported to decrease significantly (Turk & Pentland, 1991). Trying to excel, the Fischerface method (Belhumer et al., 1997) was able to obtain better error rates due to better classification and was reported to be invariant to variations in lightning. Still, the Fischerface method was more complex than Eigenfaces - resulting in longer processing time and larger storage requirements, amongst other things. The paper by Moghaddam and Pentland (1997) tried to meet the challenges of scale, but also showed relatively good performance in discarding irrelevant background and overcome variations in hair and use of glasses.

One common denominator for all the three papers mentioned in the previous paragraph are that they are used linear methods. So, although they show that their methods could overcome some variations in lightning, scale and/or pose - they are limited by the linear interpretation of the problem. Kernel PCA (Schölkopf et al.,1998) and Tensorfaces (Vasilescu and Terzopoulos, 2002) are both more “complex” methods trying to capture more advanced structures in face images to improve the results, compared to the traditional linear methods. This supports the results showing that they are better able to handle variations in lightning, scale, pose etc.

compared to the traditional linear methods (Vasilescu and Terzopoulos, 2002, Yang et al., 2002). However, both methods are reported to be relatively computationally expensive. Lastly, the method put forward by Viola and Jones (2002) stands out in terms of it's conceptual simplicity and speed. It showed good results in terms of speed, even on devices with relatively limited computational power, while still being able to handle large variations.

The quality of the work presented by the different papers can be evaluated in different ways. In terms of applicability of results Kernel PCA (Schölkopf et al.,1998) stands out as the authors try to tackle the problem of pattern recognition in general, and not just faces. However, Moghaddam and Pentland (1997) also test their probabilistic method on non-rigid objects using hands as well. Moreover, in terms of validation and testing presented in the papers, especially the paper by Moghaddam and Pentland (1997) test their method in many settings and on the widely used FERET database (NIST, 2019). However, a comparison to the Eigenface method is not presented in the paper, which would have been relevant. In general, more coherence amongst the papers in terms of which databases they use in testing would have made comparison of performance easier. Lastly, in terms of the papers' ability to explain their methods in a clear and complete way it could be argued that the papers by Viola and Jones (2002) and Belhumeur et al. (1997) stands out. Viola and Jones (2002) communicate very clearly the contributions and the intuition behind their method, while Belhumeur et al. (1997) have a very simple structure to follow.

Overall, the papers discussed in these essay all represent contributions to the advanced in research on facial recognition and detection. The earlier papers showed how one could decompose a high dimensional problem to a simpler one, while the later ones added mechanisms to enable handling nonlinear dependencies and complex structure in the image data. Lastly, especially the paper by Viola and Jones (2002) showed how several steps and different methods could be combined to significantly speed up the processing time**.**

Student ID: 10358736
COMP61342

References

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7), 711-720

Kim, K. I., Jung, K., & Kim, H. J. (2002). Face recognition using kernel principal component analysis. *IEEE signal processing letters*, *9*(2), 40-42.

Lienhart, R., & Maydt, J. (2002). An extended set of haar-like features for rapid object detection. In *Proceedings. International Conference on Image Processing* (Vol. 1, pp. I-I). IEEE.

Moghaddam, B., & Pentland, A. (1997). Probabilistic visual learning for object representation. IEEE Transactions on pattern analysis and machine intelligence, 19(7), 696-710.

NIST. (2019). *Face Recognition Technology (FERET)*. [online] Available at: https://www.nist.gov/programs-projects/face-recognition-technology-feret [Accessed 3 May 2019].

Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural computation, 10(5), 1299-1319.

Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Josa a*, *4*(3), 519-524.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1), 71-86.

Vasilescu, M. A. O., & Terzopoulos, D. (2002). Multilinear analysis of image ensembles: Tensorfaces. In *European Conference on Computer Vision* (pp. 447-460). Springer, Berlin, Heidelberg.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *CVPR (1)*, *1*, 511-518.

Yang, M. H. (2002). Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods. In *Fgr*(Vol. 2, p. 215).