An Analysis on Surrounding Literature for Detecting and Recognising Faces in Images

Darsh Jadhav
ID: 10806664

I. Summary

Face detection and recognition is an important area of technology that has many different applications, ranging from forensic investigations to authentication methods such as accessing accounts and devices. The requirements of application differs the type of implementation method required, for example, applications may require real-time detection which will require the implementation of computationally efficient algorithms without drastically increasing error rates. It is important for facial recognition technologies to be accurate as errors can make the technology unusable. For example, if the facial recognition technology in a smartphone does not accurately recognise a face, it may be unlocked by malicious individuals.

In this report, we will analyse multiple different approaches to face recognition and detection problems. We will cover the key ideas and algorithms implemented to solve the problem. Furthermore, we will compare technologies to understand the advantages and disadvantages of the implementations.

II. Overview of Literature

This section will provide an overview of each piece of literature. This will include the key ideas proposed by the paper, the technical approaches taken, and the final results produced.

A. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection by Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman [1]

Belhumeur et al. investigate and propose a facial recognition algorithm using Fisherfaces instead of correlation, Eigenfaces, or linear subspace methods. The Fisherface method intends to be "insensitive to large variation in lighting direction and facial expression" [1], resulting in lower error rates when testing on Yale and Harvard face databases.

The key ideas for this facial recognition algorithm is to use a derivation of the Fisher's Linear Discriminant (FLD) in order to find linear combinations of features to distinguish classes, referred to as Fisherfaces [1]. Belhumeur et al. mention that algorithms such as correlation (e.g. nearest neighbour), Eigenfaces, and Linear Subspaces have disadvantages for this application, deeming them to be an inferior solution for facial recognition (on the Harvard and Yale face databases) in comparison to their Fisherface implementation.

The Fisherface method performs dimensionality reduction by using linear projection, this results in lower-dimensional data that keeps its property of being linearly separable. The dimensionality reduction is done by using Principle Component Analysis (PCA), which is then followed by FLD to further reduce the dimension by c - 1, where c is the number of classes [1].

Belhumeur et al. compared the different approaches against the Fisherface method for different scenarios, these included:

- Variations in lighting
- Variations in lighting, facial expression, and eye wear
- Recognition of eye wear

The results showed that the Fisherface method had a lower error rate than the other methods outlined in all three scenarios listed above, furthermore, they found that it was less computationally expensive in comparison to other methods, even if the other methods had similar error rates in certain tests.

B. Probabilistic Visual Learning for Object Representation by Baback Moghaddam and Alex Pentland [2]

Moghaddam and Pentland propose an unsupervised learning approach for visual learning [2]. The key idea is to use the methods outlined for detecting human faces and features (such as eyes and hands), which can then be used in real world applications.

The approach taken to implement this idea is start by using eigenspace decomposition for density estimation in high-dimensional spaces [2]. Using PCA for dimensionality reduction and feature extraction, they are able to estimate complete density functions [2]. The density estimations that are produced are then used in a maximum likelihood estimation framework that is used for target detection, or classification and recognition applications.

Baback Moghaddam and Alex Pentland have tested this methodology with Bayesian and maximum a posteriori (MAP) frameworks for facial recognition. They found that this method gave them notable gains in accuracy over methods that utilised a eigenface nearest-neighbour approach.

C. Nonlinear Component Analysis as a Kernel Eigenvalue Problem by Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller [3]

Schölkopf et al. discuss the key idea of using kernels with PCA in order to implement nonlinear PCA. This method is implemented for pattern recognition applications where nonlinear extensions of PCA would be favoured over linear methods.

The kernel-based PCA algorithm works by applying a kernel function to the data-set, which is then used to compute the projections into a high-dimensional space. This is done by computing a kernel matrix, $K_{ij} = (k(x_i, x_j))_{ij}$ [3]. Kernel matrix *K* is then diagonalised, and the eigenvector expansion coefficients are normalised to return the eigenvalues which meet the requirements $\lambda_n(\alpha^n \cdot \alpha^n) = 1$, where λ is the eigenvalue. The last step is to extract the principle components from the eigenvectors that the kernels were projected onto [3].

The paper mentions that this algorithm can be used as a new method for nonlinear PCA. Applications such as feature extraction using kernel PCA can be used for classification. Schölkopf et al. mention that whilst this algorithm theoretically has its advantages, such as not having to deal with nonlinear optimisation, the choice of kernel for the problem needs to be evaluated based on the application, which is an open problem [3].

D. Eigenfaces for Recognition by Matthew Turk and Alex Pentland [4]

Matthew Turk and Alex Pentland propose a "near-real-time" computer system that is able to locate and continuously track a human's head [4]. Furthermore, this system will use characteristics of the human face that is being tracked for facial recognition. The basic idea is to produce eigenfaces via the eigenvectors produced from PCA.

The method to produce the facial recognition part of the system starts by calculating eigenfaces by taking the images of human faces and calculating the eigenvalues and their corresponding eigenvectors. Using the top *M* eigenfaces and their corresponding eigenvectors, and the normalised training data, the eigenfaces can be calculated. These eigenfaces can then be used to calculate the class vectors Ω_k, which are used to measure how similar the new target face is from the faces from the training data. If the new target face is classified correctly, then the new target face can be fed back into the training data to improve the accuracy of the model. In order to implement the motion detection element of the system, they use filtering and thresholding techniques to identify whether there is motion present. Once the system has detected motion, they use the information of where the motion is, referred to as "face space" in order to locate face [4].

The results of the developed computer system show that using the eigenface method can provide a solution for facial recognition. Their system is simple and fast, whilst giving highly accurate results. The method in which they detect facial features is an alternative solution to the active appearance model (AAM).

E. Multilinear Analysis of Image Ensembles: TensorFaces by M. Alex, O. Vasilescu, and Demetri Terzopoulos [5]

Alex et al. mention that other popular techniques for facial recognition such as the eigenface method face difficulties when main factors such as lighting, viewpoint, and facial expression are used to modify the images. Therefore, Alex et al. conduct the analysis of their use of multilinear algebra, specifically algebra of higher-order tensors for facial recognition.

Their algorithm uses tensors, also referred to as multidimensional matrix, to map sets of vector spaces [5]. These tensors undergo decomposition via an "N-Mode Single Value Decomposition (SVD)" algorithm. The SVD algorithm uses the face image ensembles to produce TensorFaces.

The results of this paper concluded that their multilinear analysis using ensemble images allows them to produce these TensorFaces that can be used to solve computer vision problems such as facial recognition.

F. Rapid Object Detection Using a Boosted Cascade of Simple Features by Paul Viola and Michael Jones [6]

The key idea of paper is to present an approach to implement an object detection system whilst attempting to balance the minimisation of computation time and maximise the object detection accuracy.

In order to do this, Viola and Jones suggest an approach that uses a combination of different algorithms. The first algorithm used is the Integral Image, which produces a representation of the input image. This algorithm computes rectangle features using an intermediate representation of the image, referred to as Integral Image [6]. These rectangle features support learning, since it provides great image representation. The next part of the approach uses an algorithm that uses a subset of important rectangle features that results in an efficient classifier. This algorithm is a variant of AdaBoost that will select the important subset of features, and train the classifier [6]. The third part of this approach utilises a cascade algorithm framework for the detection process, where multiple classifiers are used to achieve high detection rates. Each predecessor classifier that has a positive result will trigger the evaluation of its succeeding classifier. Each classifier in the cascade attempts to reduce the false positive rates and meet the optimal detection rate.

Viola and Jones create a 38 layer cascaded classifier that trained on frontal upright human faces. The detection process was efficient and fast in comparison to other implementations of face detectors.

III. Comparison of Reviewed Literature

In this section, I will compare the reviewed literature. I will dive into the strengths and weaknesses of the proposed approaches. Furthermore, I will compare these approaches to the Active Appearance Model (AAM) approach. Finally, I will attempt to rank the proposed methods based on the applicability of the results, the extent of validation and testing, and whether the description of their work is clear.

A. Strengths and Weaknesses

When examining the Fisherface approach suggested by Belhumeur et al., we can see that the Fisherface method is able to succeed when the factor of lighting variability is involved. Belhumeur et al. mention that the Fisherface method had error rates that were better than half of any other method, and the computational time was lower than the Eigenface method [1]. The Fisherface method was able to outperform the Eigenface method since the variability in lighting made it difficult for the Eigenface method to classify the images. When looking at the paper Eigenfaces for Recognition by Turk and Pentland [4], we can see that they did not consider the variability of lighting. Instead, they prioritised efficiency and balanced the detection accuracy so that they could keep their software lightweight and computationally efficient. This would potentially mean that the Eigenface method by Turk and Pentland would have a lower accuracy in comparison to the Fisherface approach implemented by Belhumeur et al. Furthermore, the Eigenface method by Turk and Pentland would suffer if there was a large variation in lighting, which is something that they did not focus on.

Viola and Jones proposed a rapid object detection using a combination of different algorithms. They utilise a cascade of classifiers to maximise the computational efficiency of the approach whilst maintaining high detection rates [6]. Their method uses a variation of AdaBoost, a machine learning technique which is known for combining weak classifiers to improve the accuracy of the overall prediction [7]. When comparing the performance of their approach, they were approximately 15 times quicker than other object detection approaches. A disadvantage of their method is how much the 3 individual classifiers rely on each other, any potential errors that are not rejected by the boosted classifiers would have adverse effects on the accuracy of the detector. Viola and Jones have mentioned that the accuracy of the detector would be improved if the individual classifiers were more independent of each other.

Alex et al. propose multilinear analysis of image ensembles to create TensorFaces [5]. They manage to use a combination of algorithms to produce these TensorFaces which are able to represent the principle axes of variation (e.g. illuminations, people, expressions, viewpoints) [5], which gives them an advantage over standard eigenfaces as these variations are all mapped. Alex et al. mention that in the future, they will pursue applying this framework to many different applications, including image recognition tasks. The disadvantage of this approach is the complexity of the approach, meaning that this approach cannot be applied to near-real-time applications, whereas Turk and Pentland's approach allows them to diversify their applications, even if the approach does not have a very high accuracy. When reviewing the paper on probabilistic visual learning for object representation by Moghaddam and Pentland, they mention their outlined method using a Bayesian/MAP technique greatly improved the accuracy of a facial recognition system which used a standard eigenface nearest-neighbour matching rule [2]. Moghaddam and Pentland examined the use of their learning technique in a couple of different applications, for example, they were able to achieve 97% accuracy in a face detection task using a set of 2000 face images.

B. Comparison of Methods to the Active Appearance Model (AAM) Approach

AAM is used to models shapes but uses the whole image rather than individual features. The AAM algorithm will generate a synthetic sample of a face, using shapes and textures, that will be used to measure the difference between the face image that is being tested on. This will then be updated to minimise the difference between the synthetic image and the target test image until the model is fully updated.

AAM is implemented using efficient search algorithms, making the less computationally expensive. Furthermore, AAM uses PCA to generate the synthetic faces by using shape-free textures. This works similarly to the implementation of Eigenfaces for Recogntion by Turk and Pentland. AAM is known for its ability to deal with variation in the image. The Fisherface algorithm proposed by Belhumeur et al. works similarly to the AAM since it also utilises textures of the image to deal with variation in illumination and facial expressions.

IV. Conclusion and Ranking Literature Papers

This section will conclude my opinions on how useful each piece of literature by ranking of each piece of literature in order of best to worst (relative to each other). Each ranking will be justified based on their generality and applicability of their results. Furthermore, I will comment on the amount of testing shown in the paper, and how clear and complete the description of the work is.

1) Probabilistic Visual Learning for Object Representation by Baback Moghaddam and Alex Pentland [2]:

- Moghaddam and Pentland have proposed a probabilistic visual learning approach for object representation. This approach is very generalised as the key algorithms outlined in this paper have been used in multiple different applications. Moghaddam and Pentland state that these approaches have been applied to the probabilistic visual modelling, detection, recognition, and coding of human faces and non-rigid objects [2].
- Moghaddam and Pentland have presented their testing in the paper, with numerous examples of how they tested their approaches, in comparison to other papers which only provide theory on the proposed algorithms.
- The way this paper is presented makes it much easier to understand in comparison to other papers.

2) Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection by Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman [1]:

- Belhumeur et al. have proposed the Fisherface method for facial recognition. This approach can be diversified to detect other objects, not limiting to faces in images.
- Belhumeur et al. have rigorously tested their approach for facial recognition, furthermore, they have tested other approaches to provide comparisons between the algorithms.
- The paper contains complex information and equations, however they have presented the work with clear explanations in order to understand the ideas they are trying to propose.

3) Eigenfaces for Recognition by Matthew Turk and Alex Pentland [4]:

- Turk and Pentland have proposed eigenfaces for facial recognition. Eigenfaces are commonly used in object recognition in images, which means they are widely applicable to different tasks.
- Turk and Pentland have tested their approach on facial recognition on both a database of face images, and real-time facial recognition. They show the tests conducting using eigenfaces which shows an appropriate amount of testing.
- The paper simplifies algorithms into steps, using descriptions of each step that makes understanding them much simpler, in comparison to the paper by Alex et al. which presents a large amount of algebra to explain the algorithms presented in their paper.

4) Rapid Object Detection Using a Boosted Cascade of Simple Features by Paul Viola and Michael Jones [6]:

- Viola and Jones propose a machine learning approach for visual object detection. This approach intends to process images rapidly whilst maintaining a high detection rate. Their approach is generalised as it can be applied to many different computer vision problems.
- Viola and Jones test their algorithms on different types of applications, furthermore, they are the first paper to be explicit on how they can improve their algorithm. This shows that they have validated their algorithm and understood the potential to improve.
- The paper explains each part of the approach well. They give examples to explain how the classifiers work, in order to simplify explanations.

5) Nonlinear Component Analysis as a Kernel Eigenvalue Problem by Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller [3]:

- Schölkopf et al. propose a nonlinear component analysis approach for feature extraction. Since this approach intends to perform feature extraction, it can be applied to any application where the features are nonlinearly related to the input variables.
- Schölkopf et al. have applied their approach to two different examples, of which only one example was applicable to a real world application (character recognition). More validation of their approaches on real world applications would increase the extent of their testing.
- This paper was went into a large amount of detail regarding the kernels which are used in this approach. The information helped distinguish differences between standard PCA and the method they are proposing.

6) Multilinear Analysis of Image Ensembles: TensorFaces by M. Alex, O. Vasilescu, and Demetri Terzopoulos [5]:

- Alex et al. propose multilinear analysis of image ensembles by producing TensorFaces. The generalisation of this method allows for improved facial recognition methods.
- Alex et al. show the use of the algorithms outlined in this paper. They presented the TensorFaces that are produced from the algorithms and explained them in great depths. However, there is no real-world applications using the results from these methods. Testing these TensorFaces would provide more information as to whether their methods produce useful results.
- This paper poorly presented the information; the algorithms were not clearly explained, but rather simply explained large portions of information with algebra, which made it much more difficult to understand the concepts in the paper.

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[2] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 696–710, 1997.

[3] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[4] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[5] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *European conference on computer vision*. Springer, 2002, pp. 447–460.

[6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.

[7] R. E. Schapire, "Explaining adaboost," in *Empirical inference*. Springer, 2013, pp. 37–52.