Can you use Viola-Jones face detection for counting people?

Samuel Jackson
Department of Computer Science
University Of Aberystwyth
Aberystwyth, Ceredigion, SY23 3FL
Email: slj11@aber.ac.uk

I. Introduction

This paper examines the Viola-Jones face detection algorithm [1] to the application of counting people in the real world. The Viola-Jones algorithm presented in [1] proposes an efficient method of performing face detection at frame rate with reasonably accurate results.

Detecting faces is a major challenge in the field of computer vision, and one which has many practical applications. Accurately determining whether a scene or photograph contains a face is provides many challenges. Yang et al. [2] lists a number of different factors affecting human face detection. Faces typically have a wide degree of variation. They have different prominent facial features, skin colour, scars, facial hair, and hair styles. The pose also causes issues. The side of a face is very different from the frontal view. Occlusions prove another major obstacle. In a crowded scene it is easy for a face to be partially or wholly excluded, glasses and sun glasses are also a common source of occlusions. The quality of the capture including the lighting, focus, and contrast can affect the image quality. The characteristics of the camera such as the lenses and sensors also affect final capture.

Another major factor which increases the difficulty of face detection is the complexity of the problem. Photographs and video generally have an incredibly large feature space to search making looking for the face at every pixel and scale an impossibly complex task to practically accomplish at frame rate.

Despite these obstacles efficient detection of faces in a scene continues to be desirable for a multitude of practical applications. One of the most common uses of a face detection system is as part of a larger system, such as a face recognition or image analysis. In order to correctly recognise a face it must first be found. In this way face detection algorithms act as a preprocessor to a larger system. Other applications include uses in image databases, social media, and video conferencing. Automatic detection of faces can be used in image management system to find faces in images as part of a precursor to automatically annotating images with meta data. The most obvious example would be the image tagging feature on sites such as Facebook. Applications to video conferencing include the need to adjust the focus of the camera relative to the speak.

The approach taken by Viola-Jones in [1], is also not restricted to detecting faces. A cascade classifier such as the one presented in [1] can be trained to detect a variety of objects other objects provided that sufficient positive and negative examples can be supplied for training. For example, work based on the Viola-Jones detection algorithm has been used to detect people [3], license plates [4], and facial features (eyes, nose, and mouth) [5].

The major contribution that the Viola-Jones algorithm makes is in the speed at which the algorithm can detect a face. Many approaches prior to Viola-Jones could offer reasonably good detection rates [6] [7] [8], but the Viola-Jones algorithm's strength is in the speed ups offered by their method.

Viola-Jones solves the problem of rapid object detection through the use of three key components. The First component is that the Haar features used to detect faces are computed using an integral image representation. This representation allows for rapid evaluation of Haar features in constant time. The second component is to use the AdaBoost algorithm to choose a small number of features that classify the training set well. The third component is to combine the weak classifiers into a cascade of classifier, with the least complex classifiers at the beginning of the cascade. This ensures that the majority of negative examples are rejected early using classifiers that are quick to evaluate, but that later, more complex nodes in the classifier minimise the occurrence of false positives. Once the cascade of classifiers has been trained it is very efficient to classify new examples by running the image through the cascade.

The rest of this paper is dedicated to reviewing the performance of Viola-Jones object detection in various circumstances and discussing its application to the titular problem. Section II reviews the what problems Viola-Jones can be used to solve and what it cannot solve. Section III discusses how the technique could be applied to the titular problem and what issues the like issues encountered would be. Section IV rounds off the the paper with conclusions drawn from the preceding sections. Finally, section V provides my self evaluation of this paper.

II. Critique of Method

Viola-Jones object detection represents a major breakthrough in the computer vision community by bringing practical speed ups in face detection through intelligent data structures and clever use of the AdaBoost algorithm to make face detection in real time a possibility. However, success of the technique is not without its limitations and the technique has several possible failure modes.

At the time of publication of the technique had results that had "detection and false positive rates which [were] equivalent to the best published results" [1] but executed in in a much smaller amount of time. Comparisons between the major approaches published at the time are given at the end of the paper, but the results are not entirely definitive due to most approaches not providing a full receiver operating characteristic (ROC) curves for their classifiers or did not publish there best results. In the case of Sung and Poggio [6] only the MIT part of the common dataset was used (as the CMU dataset did not yet exist). In spite of this, the authors show that their technique provides roughly equivalent levels of performance given the available data. The authors note that the best performance was achieved using a voting regime between three classifiers trained with different negative examples, different negative vs. positive errors, and different criteria for also positives vs. classifier size.

Unfortunately the Viola-Jones has quite a few failure modes in which it will give a high number of either false positives or false negatives. One of the biggest drawbacks of the vanilla implementation presented in the paper is that it can only detect frontal images of faces. Faces viewed for the side or in a different rotation or orientation are unlikely to be correctly detected. The technique can be used to correct detect faces as viewed from the side profile by training a different classifier with sufficient positive and negative examples. Both sides of the face could be accounted for by training a classifier for one side of the face, testing the image then flipping the image and running the classifier again to account for the other side. Similar problems occur when the classifier is presented with faces that are rotated, either in or out of the image plane. Jones and Viola [9] attempted to alleviate these issues by training multiple classifiers for each of the poses and using additional diagonal Haar like features. The choice of classifier used is decided using a decision tree. This gives a two step approach to classification in which the first step estimates the pose and the second runs selected pose classifier.

Further work by Lienhart et al. [10] provides empirical evidence that better results can be obtained from the Viola-Jones approach though the use of a number of extensions such using the gentle AdaBoost algorithm in favour of the original discrete AdaBoost implementation. They also performed experiments with a broader set of Haar features which were formed from rotated versions of the original features. These appeared to produce more accurate results despite complicating the learning procedure by successfully encoding more domain knowledge into the learning procedure that it would otherwise have difficulty learning.

More recent work published by Brubaker et al. [11] achieved improved the speed of the detector as well as reducing the training time without sacrificing accuracy.

III. Application of Method

IV. Conclusion

V. Self-Evaluation

References

[1] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[2] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 1, pp. 34–58, 2002.

[3] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 734–741.

[4] H. Zhang, W. Jia, X. He, and Q. Wu, "Learning-based license plate detection using global and local features," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 2. IEEE, 2006, pp. 1102–1105.

[5] P. I. Wilson and J. Fernandez, "Facial feature detection using haar classifiers," *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127–133, 2006.

[6]

[7] H. Schneiderman and T. Kanade, "A statistical method for 3d object detection applied to faces and cars," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 1. IEEE, 2000, pp. 746–751.

[8] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 1, pp. 23–38, 1998.

[9] M. Jones and P. Viola, "Fast multi-view face detection," *Mitsubishi Electric Research Lab TR-20003-96*, vol. 3, p. 14, 2003.

[10] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1. IEEE, 2002, pp. I–900.

[11] S. C. Brubaker, J. Wu, J. Sun, M. D. Mullin, and J. M. Rehg, "On the design of cascades of boosted ensembles for face detection," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 65–86, 2008.