An Approach for Multi-pose Face Detection Exploring Invariance by Training

Eanes Torres Pereira, Herman Martins Gomes, and João Marques de Carvalho

Universidade Federal de Campina Grande
{eanes,hmg}@computacao.ufcg.edu.br, carvalho@dee.ufcg.edu.br

Abstract. In this paper, a rotation invariant approach for face detection is proposed. The approach consists of training specific Haar cascades for ranges of in-plane face orientations, varying from coarse to fine. As the Haar features are not robust enough to cope with high in-plane rotations over many different images, they are trained only until an accented decay in precision is evident. When that happens, the range of orientations is divided up into sub-ranges, and this procedure continues until a predefined rotation range is reached. The effectiveness of the approach is evaluated on a face detection problem considering two well-known data sets: CMU-MIT [1] and FDDB [2]. When tested using CMU-MIT dataset, the proposed approach achieved accuracies higher than the traditional methods such as the ones proposed by Viola and Jones [3] and Rowley et al.[1]. The proposed approach has also achieved a large area under the ROC curve and true positive rates that were higher than the rates of all the published methods tested over the FDDB dataset.

Keywords: face detection, orientation invariance by training, adaboost, haar features, tree of classifiers.

1 Introduction

The human face is a very important way of expressing emotions and the ability to recognise them is fundamental for interpersonal social interaction and for human-computer interaction [4]. Many approaches have been proposed for face detection [5–7, 3, 8, 9], among then those which presented higher accuracies use some variation of Adaboost and weak classifiers. Although the successful application of face detection in real life situations, pose variation still remains a challenge. The approach proposed in this paper is inspired by the JointBoost method [10–12] and aims to share features of different face poses to achieve pose invariant face detection.

The Rowley et al. [6] work contains one of the first successful classifier combination approaches for rotation invariant face detection. The kernel of that detector is composed of three neural networks. The first neural network is called router, which is designed the function to determine the rotation angle of the candidate window. The router network has three layers: the first with 400 neurons (corresponding to the quantity of pixels in the image with resolution

J.F. Martínez-Trinidad et al. (Eds.): MCPR 2014, LNCS 8495, pp. 182–191, 2014.

20×20 pixels), the hidden layer has 15 neurons, and the output layer has 36 neurons (corresponding to variations from 0° to 360° augmented by 10°).

The rotation angle is classified by the router network. The face candidate is rotated to an upright position, and is given as input to two other neural networks independently trained. The result of the two neural networks is combined by a logical AND. The candidate face is only classified as face if the two networks agree in classification. Rowley et al. [6] commented on the necessity of using bootstrapping to acquire representative samples of non-face images, and to explain how they used it. The images used for testing their method are available on-line, and nowadays are commonly used to evaluate face detection approaches. This set of images is known as CMU-MIT face database.

Although the approach proposed by Rowley et al. [6] used a simple combination of classifiers, it has the drawback that all classifiers need to process every candidate window. Viola and Jones [3] proposed a method to deal with that problem. Their method combines a fast and simple feature extractor (Haar-like features [13]) with a weak classifier combination method (AdaBoost [14]).

Among the appearance based object detection approaches, the one proposed by Viola and Jones [3, 8] has achieved greater popularity and more promising results in the area. Within their framework, there are contributions to classifier training as well as to the procedure of scrutinising the image searching for objects. There followed a great number of approaches, all of them further extending that method.

For example, Huang et al. [15] proposed the training of classifiers using Sparse Granular Features and Vector Boosting. Another variation of that method is the use of Width-First-Search (WFS) to traverse the search tree. A distinctive feature of the WFS approach is that it is possible to traverse more than one path through the tree at a time. If more than one leaf is achieved at the end of the traversal, than the leaf of higher degree of confidence is used for classification. Huang et al. [15] tested their detector by using the CMU-MIT images [1] and detected correctly, for the frontal face image set, more than 97% of the faces, with less than 100 false positives.

Another variation of the Viola and Jones framework was recently proposed by Vural et al. [16]. The major contribution of the authors is the use of rotated versions of Haar-like features with angles ranging from 0° to 180°. As the authors proposed a multi-view approach, they used a combination of trained cascades to achieve such aim. In the case of face detection, they used 6 different cascades (frontal, right, left, up, down, profile) combined by neural networks to correctly classify the faces. As the used features are more powerful, lower training time is needed for training the cascades, and less features are selected to create the cascade. They tested their detector on images and videos and obtained higher results than those implemented by the Viola and Jones approach available in the OpenCV library[1], which processed images of 4 mega-pixel in 15fps. A criticism to this experimental evaluation is that they did not use any well-known face image databases to evaluate their approach, such as CMU-MIT [1] or FDDB [2].

[1] `http://opencv.willowgarage.com/wiki/`

The AdaBoost term is an acronym for adaptive boosting. It was coined by Freund and Schapire [14] to name the process of adaptively weighting classifiers in combinations such that the weak classifiers receive more attention (high weights) than the strong ones. The weights are computed based on the classification rates of each corresponding classifier, those which obtain higher hit rates are assigned lower weights, and those that achieve lower rates are assigned higher weights. Many variations of AdaBoost were proposed after the publication of the first approach. The main difference among them is the method of assigning the weights based on the classifier accuracies. For instance, the Gentle AdaBoost uses an exponential function to relate classifications with weights, and it is described by Friedman et al. [17]. Friedman et al. [17] handle the AdaBoost algorithm using an additive model $F(x) = \sum_{m=1}^{M} c_m f_m(x)$, where c_m is constant that depends on the expectation over the training data. Each f_m is a separate function for each input variable. This leads to the interpretation of each feature as a weak classifier.

Within the above context, this paper presents a method for detecting faces at different in-plane orientations with any degree of rotation in image plane; with precision rates equal or higher than those obtained by other popular approaches such as those proposed by Rowley et al. [6] and Viola and Jones [8]. However, the proposed method uses less features and is trained faster than the methods proposed by Rowley et al. [6], Viola and Jones [8], and Huang et al.[15]. An approach that shares features among different poses was conceived. The proposed approach achieves in-plane invariance by training when using cascades of Haar-like features and Adaboost. The major objective of using invariance by training is to reduce the quantity of nodes in the classifier tree, and, consequently, this reduces the detection complexity.

2 Sharing Features for Multi-pose Face Detection

The method JointBoost [10–12] may be used to share features among multiple image classes. Considering different views of the same object as different classes, one may create a multi-pose classifier by employing this method. This will be explained here as a start point for the approach proposed in this paper.

Torralba et al. [12] argue that it is possible to demonstrate, subjectively (by means of visual inspection) as well as objectively (by evaluating the features extracted from images) that some characteristics of frontal faces are present on profile faces. In the same way, characteristics of non-rotated frontal faces are present in in-plane rotated frontal faces. This observation may be extended to other object categories, such as: cars, houses, and animals. From this reasoning, Torralba et al. [12] proposed a boosting approach for multi-class problem classification, the JointBoost.

In the JointBoost approach, at every cycle of weak classifier computation, the chosen feature would be that which has the lowest classification error for the highest number of different classes. Thus, one may assert that the feature is shared among different classes. The authors say that their experiments showed

that classifiers jointly trained (using JointBoost) tend to select features that generalise well for various classes. Generally, those features are edges and blobs.

Haar-like features, such as those used by Viola and Jones [8], may also be used to generalise among multiple classes. However, such generalisation power is limited, i.e., it is not possible to obtain rotation invariance by training a classifier using only that type of features. In this paper, the term *rotation invariance by training* refers to training a Haar cascade, with GentleBoost, with rotating face images in-plane.

3 Proposed Approach

As a preliminary step of the present research, a number of experiments were performed in order to verify the possibility of training a rotation invariant face classifier simply by varying the training face rotations. Those experiments were to no avail. However, some important insights were drawn from those experiments. First, the classifier trained with Haar-like features could not generalise frontal faces in any in-plane rotation angle; however, until obtaining a number of training stages, the training converged well. As more stages are trained, the classification problem becomes more complex, and the available features cannot adequately generalise. Another idea inspired by those experiments is that classifiers obtained with invariance by training, but with a reduced set of stages, may be combined to yield a multi-pose classifier tree. Based on those observations, this section presents the proposed approach.

Figure 1 shows a simplified representation of a classifier tree obtained with invariance by training, and with reduced quantity of stages. This classifier tree may be used to detect frontal faces with any in-plane rotation. The tree root is a cascade with at most 5 stages, and it classifies frontal faces in any in-plane rotation. As the tree is binary, the orientation ranges are divided by two as the classification is propagated through the tree. The division by two allows more specific classifiers to be used in deeper tree levels. Besides, as in previous approaches, the false positives quantities go down exponentially at each tree level.

The circle on right left-hand side of the classifier tree illustrates the orientation pattern of faces. The pattern differs from the trigonometric circle, which establishes the angle of 0° corresponding to the angle of 270° presented in Figure 1. The difference between the patterns is simply an offset of 90° in such a way that the 0° corresponds to the upright face image.

Another important feature of the presented classifier tree is that each leaf actuates within a range of 20°. Consequently, the classifiers in the leafs must be trained with face images that have a variation of ±10° in relation to the angle that labels the leaf. Besides, the neighbour leafs have an intersection of 10° in their angle coverage ranges. Thus, the central angle of a leaf corresponds to the *edge* angle of the neighbour leaf.

The arrangement of angle ranges of the leafs was designed to allow redundancy and reinforcement of the classification to possibly difficult angles, e.g. cases in

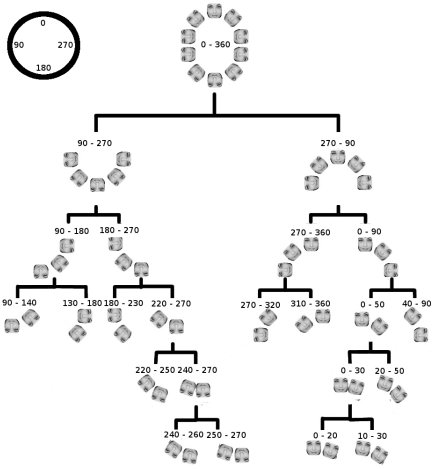


Fig. 1. Classifier tree for in-plane rotated frontal faces

which few training face samples have been used. The redundancy also allows that the candidate window be classified by more than one leaf. The leaf that will be used to classify such window will be the one that obtains the highest confidence level (which may be a threshold or a probability).

4 Experimental Evaluation

For training the cascades, 10,000 frontal face images and 10,000 profile face images were cropped from the following image databases: BioID [18], Caltech [19], CMU-PIE [20], YaleB [21] and Color FERET [22]. At each stage, 20,000 non-face image crops were used. The non-face images were selected from the author's personal images and from Naotoshi Seo web site on training Haar cascades[2]. In this section, two sets of experiments are described: experiments using the FDDB image database, and experiments using the CMU-MIT image database.

4.1 Detector Evaluation Using the FDDB Image Database

The FDDB image database (FDDB - Face Detection Data Base) [2] is a benchmark for evaluation of face detectors, without condition restrictions. That base has a companion protocol to evaluate the results obtained by the detectors applied on its images. The major motivation for the creation of the FDDB was the absence of coherent methods to compare face detectors. One image database traditionally used to evaluate face detectors is the CMU-MIT, which was used in its final version by Schneiderman and Kanade [23].

However, up to the creation of the FDDB, the image databases previously used to evaluate face detectors neither required, nor proposed an evaluation

[2] `http://note.sonots.com/SciSoftware/haartraining.html`

protocol. These bases were composed by a set of images accompanied by files containing the coordinates of the faces or some fiducial points. It must also be taken into account the fact that some face detectors are not available to the research community for evaluation. For example, one of the most popular face detectors, the detector proposed by Viola and Jones [8] has no available official implementation by the authors. A closer match of Viola and Jones' detector is implemented in OpenCV [24], which comes accompanied by some XML files with trained cascades for face detection. However, the results obtained by those detectors are inferior to the results published by Viola and Jones [8].

Apart from providing the images and the ground-truth, the FDDB provides the code that will be used to count the hits and errors. The measure used to count detection as a hit corresponds to the ratio between the area of intersection and the area of union of both the detected region and the labelled region [2]. The faces are annotated by using elliptical regions; however, the code performs the necessary conversions to compatible detection results that were marked as rectangles. The authors presented as features of their base: (1) the great quantity of images and faces: 2845 and 5171, respectively; (2) a great range of difficulties (occlusions, poses, low resolution, and faces with bad focus); (3) the specification of face regions using elliptical regions.

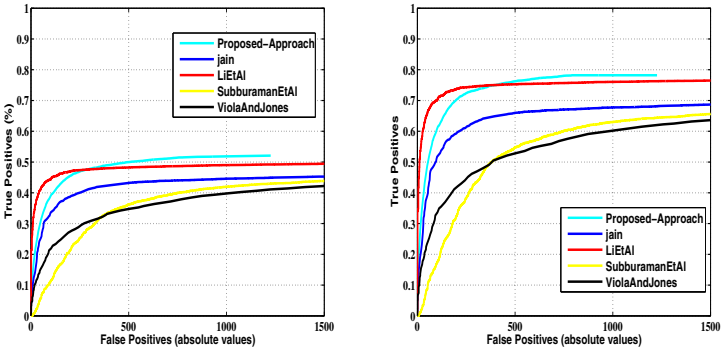
There are two ways to evaluate detectors using the FDDB: 10-fold cross-validation and training without restrictions. In the first case, the cumulative performance is reported as a mean curve of the 10 ROC curves (Receiver Operating Characteristic). In the second case, one is allowed to use images that are not part of the base to train the classifiers. However, in this case, the set is also divided into 10 parts, and the resulting ROC curve is obtained from the average of curves. The experimental mode was applied without restrictions in order to evaluate the proposed method. At the time this paper was written, all results presented on the FDDB site did use such experimentation mode.

Another peculiarity of the FDDB is that there are two evaluation metrics: the discrete metric and the continuous metric. The discrete metric counts as hit at every detection at which the ratio between the area of intersection and area of union with the ground-truth region is higher than 0.5. The continuous metric assigns a score to the detection equivalent to the ratio between areas of intersection and union.

On the result page of the FDDB's site[3] there are ROC curves for face detectors that had published papers and results without announced publications. The graphics of the ROC curves are separately presented as: one graphic pair (with continuous and discrete metrics, respectively) for the detectors whose methods were published, and a pair of graphics for the detectors whose approaches were not published. Among the detectors with published methods, those with higher results were proposed by Li et al. [25]. Other published results used for comparison are the results obtained by Jain and Miller [26], Subburaman and Marcel [27]. The OpenCV implementation represents the Viola and Jones approach [3].

[3] `http://vis-www.cs.umass.edu/fddb/results.html`

In Figures 2(a) and 2(b), detection results are presented for the two metrics. The curves labelled as *LiEtAl* refer to the results with the highest results found in the FDDB's site.



(a) Continuous metric. (b) Discrete metric.

Fig. 2. Comparison of face detection results in the FDDB database

The results showed in Figures 2(a) and 2(b) were obtained by using the following methods: upright frontal detector trained via the proposed approach, the Jain and Miller detector [26] (which are the authors of FDDB), the Li et al. [25] detector, the Subburaman and Marcel [27] detector, and by the Viola and Jones detector represented by OpenCV implementation [24]. The small areas under the ROC curves are due to the fact that the base is composed by images of faces in different poses, and the detectors were trained only for frontal face detection with extreme in-plane rotations.

4.2 Detector Evaluation Using the CMU-MIT Image Database

To the best of our knowledge, the image database that has been largely used for rotation invariant face detector evaluation is the CMU-MIT [6]. The FDDB database has variations in-plane and out-of-plane, but does not present extreme variations (e.g., up-side down faces). Additionally, the CMU-MIT has been mostly used to compare results of diverse approaches. Figure 3 presents the results of face detection by using the rotated set of CMU-MIT images for four detectors: the proposed approach, the Jones and Viola's detector [28], the Rowley et al.'s detector [6], and the Huang et al.'s detector [15]. The curves which represent the results obtained by Jones and Viola [28] and Rowley et al. [6] were constructed from the tables of results reported in their corresponding papers.

The curve representing the results of Rowley et al. [6] has a sawtooth shape as it was obtained by the interpolation of just four points. The hit rate verification metric used in that approach is similar to the one mentioned by Lienhart

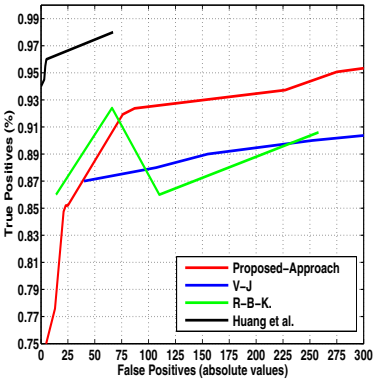


Fig. 3. Face detection results in the CMU-MIT rotated images (rotated set) database

et al. [29]. The results obtained by Huang et al. [15] are very high, with true positive rates higher than 90% without the occurrence of any false positive. A possible criticism to those results is related to its range of variation: Why did not the authors vary the parameters of testing sufficiently to show the results for quantities of false positives higher than 100? Another fact to be questioned is that Huang et al. [15] did not mention which metric was used to measure the hits, and they did not comment on the use of distances between the centers of the regions or on the use of the face areas.

In terms of complexity, the algorithms that use tree of classifiers may be compared by the quantity of stages. The detector created by Jones and Viola [28] is composed of two cascades: one to estimate rotation (with 11 stages), and another to distinguish between face and non-face (35 stages). Thus, according to this approach, a face candidate must pass through 46 stages of evaluation to be classified as face.

The face detector proposed by Huang et al. [15] has 234 nodes (each node corresponds to a weak classifier) and 18 stages. The classifier tree proposed in this paper has 192 nodes and 6 stages. It has 64 nodes for each view: frontal, left profile, and right profile. Thus, a candidate window should pass through less nodes, and less classification stages, when submitted to the detector proposed in this paper. Besides, the range of rotation angles used by the proposed detector are better fine-grained ($\pm 10°$) in relation to the range of Huang et al. [15]($\pm 15°$), that allows for much higher precision when estimating rotation angle.

Huang et al.[15] did not explain why they did not evaluate their detector using the CMU-MIT image test set without extreme rotations (Test Sets A, B, and C). Possibly, their detector would present inferior results, because it was trained with face images with resolution of 24×24 pixels, and it is known the mentioned image set has many images with lower resolution.

5 Conclusion

In this paper, a new approach for rotation invariant object detection is proposed. The major feature of the proposed approach is the rotation by training, in which sets of rotated images are presented in the classifier training stage and the features that best describe that set of rotated images are selected for usage as part of weak classifiers. The proposed approach demonstrates the viability of using weak and non-invariant features in order to obtain a robust and rotation invariant object detector. This is possible due to the fact that Haar-like features exhibit some degree of generalisation among multiple classes that may be exploited using classifier trees as it was explained in Section 2. Two well-known image databases were used for experimental evaluation: CMU-MIT, and FDDB. The proposed approach was evaluated in a face detection scenario and obtained better results than all the other published approaches evaluated on FDDB image database. The proposed approach obtained higher precisions than those of Rowley et al. [6] and Jones and Viola [28].

References

1. Rowley, H., Baluja, S., Kanade, T.: Neural network-basead face detection. IEEE Trans. Pattern Anal. Machine Intell. 20, 23–38 (1998)
2. Jain, V., Learned-Miller, E.: Fddb: A benchmark for face detection in unconstrained settings. Technical report, University of Massachusetts, Amherst (2010) (Relatório Técnico UM-CS-2010-009)
3. Viola, P., Jones, M.: Robust real-time object detection. In: Second Int. Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing, and Sampling, pp. 1–25 (2001)
4. Gong, S., Xiang, T.: Visual Analysis of Behaviour: From Pixels to Semantics. Springer (2011)
5. Sung, K.K., Poggio, T.: Example-based learning for view-based human face detection. IEEE Trans. Pattern Anal. Machine Intell. 20(1), 39–51 (1998)
6. Rowley, H., Baluja, S., Kanade, T.: Rotation invariant neural network-based face detection. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 38–44 (1998)
7. Schneiderman, H., Kanade, T.: A statistical model for 3d object detection applied to faces and cars. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–6 (2000)
8. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. of Comp. Vis. 57(2), 137–154 (2004)
9. Chen, H.Y., Huang, C.L., Fu, C.M.: Hybrid-boost learning for multi-pose face detection and facial expression recognition. In: IEEE International Conference on Multimedia and Expo., pp. 671–674 (2007)
10. Torralba, A., Murphy, K.P., Freeman, W.T.: Sharing features: efficient boosting procedures for multiclass object detection. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition - CVPR, pp. 762–769 (2004)
11. Torralba, A., Murphy, K.P., Freeman, W.T.: Shared features for multiclass object detection. In: Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (eds.) Toward Category-Level Object Recognition. LNCS, vol. 4170, pp. 345–361. Springer, Heidelberg (2006)

12. Torralba, A., Murphy, K.P., Freeman, W.T.: Sharing visual features for multiclass and multiview object detection. IEEE Trans. Pattern Anal. Machine Intell. 29(5), 854–869 (2007)
13. Papageorgiou, C.P., Oren, M., Poggio, T.: A general framework for object detection. In: Sixth Int. Conf. on Computer Vision, pp. 555–562 (1998)
14. Freund, Y., Schapire, R.E.: A short introduction to boosting. J. of Japanese Soc. for Artif. Intell. 5(14), 771–780 (1999)
15. Huang, C., Ai, H., Li, Y., Lao, S.: High-performance rotation invariant multiview face detection. IEEE Trans. Pattern Anal. Machine Intell. 29(4), 671–686 (2007)
16. Vural, S., Mae, Y., Uvet, H., Arai, T.: Multi-view fast object detection by using extended haar filters in uncontrolled environments. Patt. Recog. Lett. 33(2), 126–133 (2012)
17. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: A statistical view of boosting. The Annals of Statistics 28(2), 337–407 (2000)
18. Jesorsky, O., Kirchberg, K.J., Frischholz, R.W.: Robust face detection using the hausdorff distance. In: Bigun, J., Smeraldi, F. (eds.) AVBPA 2001. LNCS, vol. 2091, p. 90. Springer, Heidelberg (2001)
19. Weber, M.: Frontal face dataset (2010), `http://www.vision.caltech.edu/html-files/archive.html`
20. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. IEEE Trans. Pattern Anal. Machine Intell. 25(12), 1615–1618 (2003)
21. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Machine Intell. 23(6), 643–660 (2001)
22. Philips, P.J., Moon, H.: The feret evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Machine Intell. 22(10), 1090–1104 (2000)
23. Schneiderman, H., Kanade, T.: A statistical method for 3d object detection applied to faces and cars. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 746–751 (2000)
24. Bradsky, G., Kaehler, A.: Learning OpenCV: Computer Vision with the OpenCV Library. O'Really (2008)
25. Li, J., Wang, T., Zhang, Y.: Face detection using surf cascade. In: IEEE International Conference on Computer Vision - ICCV, pp. 2183–2190 (2011)
26. Jain, V., Learned-Miller, E.: Online domain adaption of a pre-trained cascade of classifiers. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 577–584 (2011)
27. Venkatesh, B.S., Marcel, S.: Fast bounding box estimation based face detection. In: European Conf. on Computer Vision (ECCV) - Workshop on Face Detection, pp. 1–14 (2010)
28. Jones, M., Viola, P.: Fast multi-view face detection. Technical report, Mitsubishi Electric Research Laboratories, Technical Report TR2003-96 (2003)
29. Lienhart, R., Kuranov, A., Pisarevsky, V.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. Technical report, Micropocessor Research Lab and Intel Labs (2002)