Face Detection

Ming-Hsuan Yang

University of California, Merced, CA 95344
mhyang@ucmerced.edu

Synonyms

Detecting faces

Definition

Face detection is concerned with finding whether or not there are any faces in a given image (usually in gray scale) and, if present, return the image location and content of each face. This is the first step of any fully automatic system that analyzes the information contained in faces (e.g., identity, gender, expression, age, race and pose). While earlier work dealt mainly with upright frontal faces, several systems have been developed that are able to detect faces fairly accurately with in-plane or out-of-plane rotations in real time. Although a face detection module is typically designed to deal with single images, its performance can be further improved if video stream is available.

Main Body Text

Introduction

The advances of computing technology has facilitated the development of real-time vision modules that interact with humans in recent years. Examples abound, particularly in biometrics and human computer interaction as the information contained in faces needs to be analyzed for systems to react accordingly. For biometric systems that use faces as non-intrusive input modules, it is imperative to locate faces in a scene before any recognition algorithm can be applied. An intelligent vision-based user interface should be able to tell the attention focus of the user (i.e., where the user is looking at) in order to respond accordingly. To detect facial features accurately for applications such as digital cosmetics, faces need to be located and registered first to facilitate further processing. It is evident that face detection plays an important and critical role for the success of any face processing systems.

The face detection problem is challenging as it needs to account for all possible appearance variation caused by change in illumination, facial features, occlusions, etc. In addition, it has to detect faces that appear at different scale, pose, with in-plane rotations. In spite of all these difficulties, tremendous progress has been made in the last decade and many systems have shown impressive real-time performance. The recent advances of these algorithms have also made significant contributions in detecting other objects such as humans/pedestrians, and cars.

Operation of a Face Detection System

Most detection systems carry out the task by extracting certain properties (e.g., local features or holistic intensity patterns) of a set of training images acquired at a fixed pose (e.g., upright frontal pose) in an off-line setting. To reduce the effects of illumination change, these images are processed with histogram equalization [3, 1] or standardization (i.e., zero mean unit variance) [2]. Based on the extracted properties, these systems typically scan through the entire image at every possible

location and scale in order to locate faces. The extracted properties can be either manually coded (with human knowledge) or learned from a set of data as adopted in the recent systems that have demonstrated impressive results [3, 1, 4, 5, 2]. In order to detect faces at different scale, the detection process is usually repeated to a pyramid of images whose resolution are reduced by a certain factor (e.g., 1.2) from the original one [3, 1]. Such procedures may be expedited when other visual cues can be accurately incorporated (e.g., color and motion) as pre-processing steps to reduce the search space [5]. As faces are often detected across scale, the raw detected faces are usually further processed to combine overlapped results and remove false positives with heuristics (e.g., faces typically do not overlap in images) [1] or further processing (e.g., edge detection and intensity variance).

Numerous representations have been proposed for face detection, including pixel-based [3, 1, 5], parts-based [6, 4, 7], local edge features [8, 9], Haar wavelets [10, 4] and Haar-like features [2, 11]. While earlier holistic representation schemes are able to detect faces [3, 1, 5], the recent systems with Haar-like features [2, 12, 13] have demonstrated impressive empirical results in detecting faces under occlusion. A large and representative training set of face images is essential for the success of learning-based face detectors. From the set of collected data, more positive examples can be synthetically generated by perturbing, mirroring, rotating and scaling the original face images [3, 1]. On the other hand, it is relatively easier to collect negative examples by randomly sampling images without face images [3, 1].

As face detection can be mainly formulated as a pattern recognition problem, numerous algorithms have been proposed to learn their generic templates (e.g., eigenface and statistical distribution) or discriminant classifiers (e.g., neural networks, Fisher linear discriminant, sparse network of Winnows, decision tree, Bayes classifiers, support vector machines, and AdaBoost). Typically, a good face detection system needs to be trained with several iterations. One common method to further improve the system is to bootstrap a trained face detector with test sets, and re-train the system with the false positive as well as negatives [1]. This process is repeated several times in order to further improve the performance of a face detector. A survey on these topics can be found in [5], and the most recent advances are discussed in the next section.

Recent Advances

The AdaBoost-based face detector by Viola and Jones [2] demonstrated that faces can be fairly reliably detected in real-time (i.e., more than 15 frames per second on 320 by 240 images with desktop computers) under partial occlusion. While Haar wavelets were used in [10] for representing faces and pedestrians, they proposed the use of Haar-like features which can be computed efficiently with integral image [2]. Figure 1 shows four types of Haar-like features that are used to encode the horizontal, vertical and diagonal intensity information of face images at different position and scale. Given a sample image of 24 by 24 pixels, the exhaustive set of parameterized Haar-like features (at different position and scale) is very large (about 160,000). Contrary to most of the prior algorithms that use one single strong classifier (e.g., neural networks and support vector machines), they used an ensemble of weak classifiers where each one is constructed by thresholding of one Haar-like feature. The weak classifiers are selected and weighted using the AdaBoost algorithm [14]. As there are large number of weak classifiers, they presented a method to rank these classifiers into several cascades using a set of optimization criteria. Within each stage, an ensemble of several weak classifiers is trained using the AdaBoost algorithm. The motivation behind the cascade of classifier is that simple classifiers at early stage can filter out most negative examples efficiently, and stronger classifiers at later stage are only necessary to deal with instances that look like faces. The final detector, a 38 layer cascade of classifiers with 6,060 Haar-like features, demonstrated impressive real-time performance with fairly high detection and low false positive rates. Several extensions to detect faces in multiple views with in-plane ration have since been proposed [15, 12, 13]. An implementation of the AdaBoost-based face detector [2] can be found in the Intel OpenCV library.

Despite the excellent run-time performance of boosted cascade classifier [2], the training time of such a system is rather lengthy. In addition, the classifier cascade is an example of degenerate decision tree with an unbalanced data set (i.e., a small set of positive examples and a huge set of negative ones). Numerous algorithms have been proposed to address these issues and extended to detect faces in multiple views. To handle the asymmetry between the positive and negative data sets, Viola and Jones proposed the asymmetric AdaBoost algorithm [16] which keeps most of the weights on the the positive examples. In [2], the AdaBoost algorithm is used to select a specified number of weak classifiers with lowest error rates for each cascade and the process is repeated until a set of optimization criteria (i.e., the number of stages, the number of features of each stage, and the detection/false positive rates) is satisfied. As each weak classifier is made of one single Haar-like feature, the process within each stage can be considered as a feature selection problem. Instead of repeating the feature selection process at each stage, Wu et al. [17] presented a greedy algorithm for determining the set of features for all stages first before training the cascade classifier. With the greedy feature selection algorithm used as a pre-computing procedure, they reported that the training time of the classifier cascade with AdaBoost is reduced by 50 to 100 times. For learning in each stage (or node)

within the classifier cascade, they also exploited the asymmetry between positive and negative data using a linear classifier with the assumptions that they can be modeled with Gaussian distributions [17]. The merits and drawbacks of the proposed linear asymmetric classifier as well as the classic Fisher linear discriminant were also examined in their work. Recently, Pham and Cham proposed an online algorithm that learns asymmetric boosted classifiers [18] with significant gain in training time.

In [19], an algorithm that aims to automatically determine the number of classifiers and stages for constructing a boosted ensemble was proposed. While a greedy optimization algorithm was employed in [2], Brubaker et al. proposed an algorithm for determining the number of weak classifiers and training each node classifier of a cascade by selecting operating points within a receiver operator characteristic (ROC) curve [20]. The solved the optimization problem using linear programs that maximize the detection rates while satisfying the constraints of false positive rates [19].

Although the original four types of Haar-like features are sufficient to encode upright frontal face images, other types of features are essential to represent more complex patterns (e.g., faces in different pose) [15, 12, 13, 11, 21]. Most systems take a divide-and-conquer strategy and a face detector is constructed for a fixed pose, thereby covering a wide range of angles (e.g., yaw and pitch angles). A test image is either sent to all detectors for evaluation, or to a decision module with a coarse pose estimator for selecting the appropriate trees for further processing. The ensuing problems are how the types of features are constructed, and how the most important ones from a large feature space are selected. More generalized Haar-like features are defined in [12, 11] in which the rectangular image regions are not necessarily adjacent, and furthermore the number of such rectangular blocks is randomly varied [11]. Several greedy algorithms have been proposed to select features efficiently by exploiting the statistics of features before training boosted cascade classifiers [17, 21].

There are also other fast face detection methods that demonstrate promising results, including the component-based face detector using Naive Bayes classifiers [4], the face detectors using support vector machines [22, 23, 7], the Anti-face method [24] which consists of a series of detectors trained with positive images only, and the energy-based method [25] that simultaneously detects faces and estimates their pose in real time.

Quantifying Performance

There are numerous metrics to gauge the performance of face detection systems, ranging from detection frame rate, false positive/negative rate, number of classifier, number of feature, number of training image, training time, accuracy and memory requirements. In addition, the reported performance also depends on the definition of a "correct" detection result [1, 5]. Figure 2 shows the effects of detection results versus different criteria, and more discussions can be found in [1, 5]. The most commonly adopted method is to plot the ROC curve using the de facto standard MIT + CMU data set [1] which contains frontal face images. Another data set from CMU contains images with faces that vary in pose from frontal to side view [4]. It has been noticed that although the face detection methods nowadays have impressive real-time performance, there is still much room for improvement in terms of accuracy. The detected faces returned by state-of-the-art algorithms are often a few pixels (around 5) off the "accurate" locations, which is significant as face images are usually standardized to 21 by 21 pixels. While such results are the trade-offs between speed, robustness and accuracy, they inevitably degrade the performance of any biometric applications using the contents of detected faces. Several post-processing algorithms have been proposed to better locate faces and extract facial features (when the image resolution of the detected faces is sufficiently high) [26, 27].

Applications

As face detection is the first step of any face processing system, it finds numerous applications in face recognition, face tracking, facial expression recognition, facial feature extraction, gender classification, clustering, attentive user interfaces, digital cosmetics, biometric systems, to name a few. In addition, most of the face detection algorithms can be extended to recognize other objects such as cars, humans, pedestrians, and signs, etc [5].

Summary

Recent advances in face detection have created a lot of exciting and reasonably robust applications. As most of the developed algorithms can also be applied to other problem domains, it has broader impact than detecting faces in images alone. Future research will focus on improvement of detection precision (in terms of location), online training of such detectors, and novel applications.

Related Entries

Biometric Algorithm, Ensemble Learning, Face Recognition (Systems), Face Tracking, Facial Expression Recognition, Machine Learning, Supervised Learning, Surveillance.

References

1. Rowley, H., Baluja, S., Kanade, T.: Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(1) (1998) 23–38
2. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision **57**(2) (2004) 137–154
3. Sung, K.K., Poggio, T.: Example-based learning for view-based human face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(1) (1998) 39–51
4. Schneiderman, H., Kanade, T.: Object detection using the statistics of parts. International Journal of Computer Vision **56**(3) (2004) 151–177
5. Yang, M.H., Kriegman, D., Ahuja, N.: Detecting faces in images: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(1) (2002) 34–58
6. Mohan, A., Papageorgiou, C., Poggio, T.: Example-based object detection in images by components. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(4) (2001) 349–361
7. Heisele, B., Serre, T., Poggio, T.: A component-based framework for face detection and identification. International Journal of Computer Vision **74**(2) (2007) 167–181
8. Amit, Y., Geman, D.: A computational model for visual selection. Neural Computation **11**(7) (1999) 1691–1715
9. Fleuret, F., Geman, D.: Coarse-to-fine face detection. International Journal of Computer Vision **41**(12) (2001) 85–107
10. Papageorgiou, C., Poggio, T.: A trainable system for object recognition. International Journal of Computer Vision **38**(1) (2000) 15–33
11. Dollar, P., Tu, Z., Tao, H., Belongie, S.: Feature mining for image classification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2007)
12. Li, S., Zhang, Z.: Floatboost learning and statistical face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(9) (2004) 1112–1123
13. Huang, C., Ai, H., Li, Y., Lao, S.: High-performance rotation invariant multiview face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(4) (2007) 671–686
14. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and application to boosting. Journal of computer and system sciences **55**(1) (1997) 119–139
15. Jones, M., Viola, P.: Fast multi-view face detection. Technical Report TR2003-96, Mitsubishi Electrical Research Laboratories (2003)
16. Viola, P., Jones, M.: Fast and robust classification using asymmetric Adaboost and a detector cascade. In: Advances in Neural Information Processing Systems. (2002) 1311–1318
17. Wu, J., Brubaker, S.C., Mullin, M., Rehg, J.: Fast asymmetric learning for cascade face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(3) (2008) 369–382
18. Pham, M.T., Cham, T.J.: Online learning asymmetric boosted classifiers for object detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2007)
19. Brubaker, S.C., Wu, J., Sun, J., Mullin, M., Rehg, J.: On the design of cascades of boosted ensembles for face detection. International Journal of Computer Vision **77**(1-3) (2008) 65–86
20. Provost, F., Fawcett, T.: Robust classification for imprecise environments. Machine Learning **42**(3) (2001) 203–231
21. Pham, M.T., Cham, T.J.: Fast training and selection and Haar features using statistics in boosting-based face detection. In: Proceedings of IEEE International Conference on Computer Vision. (2007)
22. Oren, M., Papageorgiou, C., Sinha, P., Osuna, E., Poggio, T.: Pedestrian detection using wavelet templates. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (1997) 193–199
23. Romdhani, S., Torr, P., Schölkopf, B., Blake, A.: Computationally efficient face detection. In: Proceedings of the Eighth IEEE International Conference on Computer Vision. Volume 2. (2001) 695–700
24. Keren, D., Osadchy, M., Gotsman, C.: Antifaces: A novel fast method for image detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(7) (2001) 747–761
25. Osadchy, M., LeCun, Y., Miller, M.: Synergistic face detection and pose estimation with energy-based models. Journal of Machine Learning Research (2007) 1197–1214
26. Moriyama, T., Kanade, T., Xiao, J., Cohn, J.: Meticulously detailed eye region model and its application to analysis of facial images. IEEE Transactions on Pattern Analysis and Machine Intelligence **5**(28) (2006) 73800752
27. Ding, L., Martinez, A.: Precise detailed detection of faces and facial features. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2008)
28. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). The Annals of Statistics **28**(2) (2000) 337–407

Definitional Entries

AdaBoost

AdaBoost (short for Adaptive Boosting) is a machine learning algorithm formulated by Freund and Schapire [14] that learns a strong classifier by combining an ensemble of weak (moderately accurate) classifiers with weights. The discrete AdaBoost algorithm was originally developed for classification using the exponential loss function and is an instance within the boosting family. See [28] for deriving boosting algorithm from the perspective of function approximation with gradient descent and applications for regression.

Haar-like features

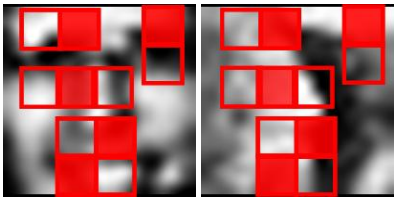
Similar to the what Haar wavelets are developed for basis functions to encode signals, the objective of two-dimensional Haar features is to collect local oriented intensity difference at different scale for representing image patters. This representation transforms an image from pixel space to the space of wavelet coefficients with an over-complete dictionary of features. See [10] for how such features can be used to represent face and pedestrians images. The Haar-like features, similar to Haar wavelets, compute local oriented intensity difference using rectangular blocks (rather than pixels) which can be computed efficiently with the integral image [2].

ROC curve

An ROC (receiver operating characteristic) curve is a plot commonly used in machine learning and data mining for exhibiting the performance of a classifier under different criteria. The y-axis is the true positive and the x-axis is the false positive (i.e., false alarm). A point on ROC curve shows that the trade-off between the achieved true positive detection rate and the accepted false positive rate. See [20] for details.

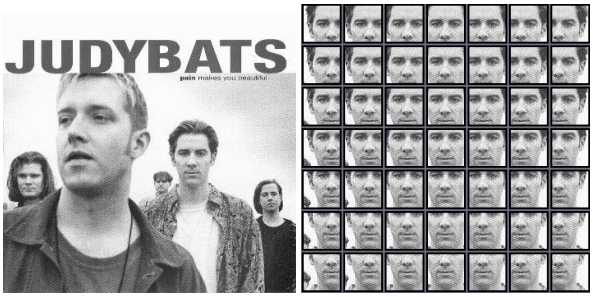
Classifier cascade

In face detection, a classifier cascade is a degenerate decision tree where each node (decision stump) consists of a binary classifier. In [2], each node is a boosted classifier consisting of several weak classifiers. These boosted classifiers are constructed so that the ones near the root can be computed very efficiently at very high detection rate with acceptable false positive rate. Typically, most patches in a test image can be classified as faces/non-faces using simple classifiers near the root, and relatively few difficult ones need to be analyzed by nodes with deeper depth. With this cascade structure, the total computation of examining all scanned image patches can be reduced significantly.



(a) Face images (b) Non-face images

Fig. 1. Four types of Haar-like features. These features appear at different position and scale. The Haar-like features are computed as the difference of dark and light regions. They can be considered as features that collect local edge information at different orientation and scale. The set of Haar-like features is large, and only a small amount of them are learned from positive and negative examples for face detection.



(a) Test image (b) Detection results

Fig. 2. Detection results depend heavily on the adopted criteria. Suppose all the sub-images in (b) are returned as face patterns by a detector. A loose criterion may declare all the faces as "successful" detections while a more strict one would declare most of them as non-faces.