2nd International Conference on Automated Face and Gesture Recognition

Detection of Human Faces Using Decision Trees

Jeffrey Huang, Srinivas Gutta, and Harry Wechsler
Department of Computer Science
George Mason University
Fairfax, VA 22030

{rhuang, sgutta, wechsler}@cs.gmu.edu
http://chagall.gmu.edu/FORENSIC

Abstract

This paper proposes a novel algorithm for face detection using decision trees (DT) and shows its generality and feasibility using a data base consisting of 2,340 face images from the FERET data base (corresponding to 817 subjects and including 190 sets of duplicates) over a semi uniform background. The approach used for face detection involves three main stages, those of location, cropping, and post processing. The first stage finds a rough approximation for the possible location of the face box, the second stage will refine it, and the last stage decides whether a face is present in the image and if the answer is positive would normalize the face image. The algorithm does not require multiple (scale) templates and the accuracy achieved is 96%. Accuracy is based on the visual observation that the face box includes both eyes, nose, and mouth, and that the top side of the box is below the hairline. Experiments were also performed to assess the accuracy of the algorithm in rejecting images where no face is present. Using a small data base of 25 images of various but complex backgrounds the algorithm failed on two images for an overall accuracy rate of 92%.

Key words: face detection, face normalization, histogram equalization, decision tree (DT), face recognition, FERET, projection profiles

1 Introduction

People can easily recognize familiar human faces. Automated face recognition, however, is a difficult problem since face images can vary considerably in terms of facial expressions, age, image quality and photometry, geometry, occlusion, and disguise (Samal and Iyengar, 1992; Chellapa et. al., 1995). Face recognition usually starts with (the location and) detection of face patterns, proceeds by normalizing the face images to account for geometrical and illumination changes, possibly using information about the facial landmarks, identifies the faces using appropriate classification algorithms, and postprocesses the results using model-based schemes and logistic feedback. An automated system that performs these operations will find countless (security) applications.

This paper is concerned with the face detection task. The algorithm decides first if a face is present, and if the face is present it crops ('box') the face. The reason for face detection is to focus computational resources on the face area (and thus speed up further processing), and to provide the spatial context for the normalization stage. The objective of any face detection system is to segment the face (foreground) from its background and provide its 'box' boundary.

Most face recognition systems assume that the face box is already available. Amongst those trying to detect the face box, Yang and Huang (1993) have developed a face detection system for complex backgrounds using a hierarchical knowledge-based method. The accuracy reported is 83% on a database of 60 subjects. Burel and Carel (1994) use a multi-resolution detection and localization phase consisting of scanning each image at seven scales. For each scale, the window contents are normalized to a standard size, and then propagated through a Multi Layer Perceptron (MLP). The accuracy reported is 85% on a database of 40 subjects. As another example, Rowley et. al. (1995) use a cluster of back propagation networks for face detection. A (retinal) connected neural network examines small windows of an image, and decides whether each window contains a face. The system arbitrates between multiple networks to improve performance over a single network. The network is trained using bootstrapping and the accuracy reported is 92.9 % on a data base of 65 images from complex backgrounds. From the publications mentioned above it is clear that the size of the image data base is quite restricted and that all the image data is used for training. As a consequence no conclusions can be drawn about the ability of such methods to generalize (on unseen test data) and to scale up on large image data bases, possibly consisting of several thousands images.

This paper proposes a novel algorithm for face detection using decision trees (DT) and shows its generality and feasibility using a data base consisting of 2,340 face images from the FERET data base (corresponding to 817 subjects and including 190 sets of duplicates) over a semi uniform background.

The algorithm used does not require multiple (scale) templates and the accuracy achieved is 96%. Accuracy regarding cropping the face is based on the visual observation that the box includes both eyes, nose, and mouth, and that the top side of the box is below the hairline. Experiments were also performed to assess the accuracy of the algorithm in rejecting images where no face is present. Using a small data base of 25 images of various but complex backgrounds the algorithm failed on two images for an overall accuracy rate of 92%.

2. Decision Trees

The basic aim of any concept-learning symbolic system is to construct rules for classifying objects given a *training set* of objects whose class labels are known. The objects are described by a fixed collection of attributes, each with its own set of discrete values and each object belongs to one of two classes. The rules derived in our case will form a decision tree (DT). The decision trees are derived using C4.5, the most commonly used algorithm for the induction of decision trees (Quinlan, 1986). The C4.5 algorithm uses an information-theoretical approach, the entropy, for building the decision tree. In an iterative fashion C4.5 determines the feature which is most discriminatory and then it dichotomizes (splits) the data into classes categorized by this feature. The next significant feature of each of the subsets is then used to further partition them and the process is repeated recursively until each of the subsets contain only one kind of labeled data. The resulting structure is called a decision tree, where nodes stand for feature discrimination tests while their exit branches stand for those subclasses of labeled examples satisfying the test. An unknown example is classified by starting at the root of the tree, performing the sequential tests and following the corresponding branches until a leaf (terminal node) is reached indicating that some class has been decided on. Decision trees are disjunctive, since each branch leaving a decision node corresponds to a separate disjunctive case. After decision trees are constructed a tree pruning mechanism is invoked. Pruning is used to reduce the effect of noise in the learning data. It discards some of the unimportant sub-trees and retains those covering the largest number of examples. The tree obtained thus provide a more general description of the learned concept.

3. Face Detection

The first stage finds a rough approximation for the possible location of the face box, the second stage will refine it, and the last stage decides whether a face is present in the image and if the answer is positive would normalize the face image. The first stage locates BOX1, possibly surrounding the face, using simple but fast algorithms in order to focus processing for the cropping stage. The location stage consists of three steps: (**i**) histogram equalization, (**ii**) edge detection, and (**iii**) analysis of projection profiles. The cropping stage labels each 8 x 8 window from BOX1 as face or non-face using decision trees (DT). The DT are induced ('learned') from both positive ('face') and negative ('non-face') examples expressed in terms of features such as entropy, mean, and standard deviation (sd). The labeled output of the cropping stage, is post processed to (**i**) decide if a face is present, and if it is present to (**ii**) derive its BOX2 in terms of straight boundaries, and to (**iii**) normalize the box to yield uniform sized boxes BOX3 for all face images. The architecture for implementing the above procedure for detecting the face box is shown below in Fig. 1.

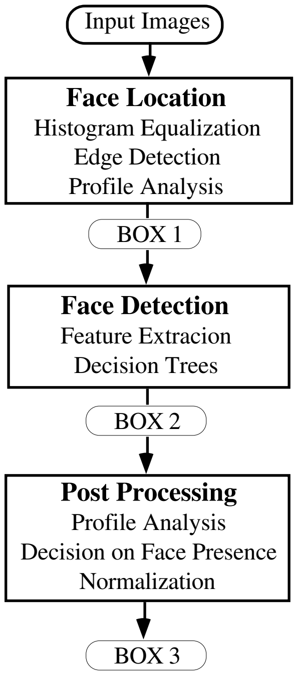

Figure 1. Face Detection

3.1 Location

The location stage indicates where is most likely to find the face contents and thus focuses attention for the next stage, that of (face) cropping. The whole procedure is straightforward and yields a rough approximation for the sought after face box.

Histogram equalization makes the image's cumulative histogram approximately linear and accounts for illumination differences as experienced during the image acquisition process. Edge detection finds transitions ('edges') in the intensity image using the Marr-Hildreth operator [Hildreth and Marr,1980]. The location stage proceeds now with horizontal and vertical projections along the x - and y - axes, respectively. The resulting profiles are searched for their maximum and minimum values using model-based ('face') constraints. The horizontal projection profile is searched for its extrema points by first detecting all its local extrema (**min** and **max**) points, and then measuring the distance **d** between each **min** and the following **max** peak. The top side of the boundary box is placed at the **max** peak location yielding the largest **d** distance. A similar procedure is used to analyze the vertical projection and yields the left and right sides of the boundary box. BOX1 is finally completed on its bottom side such that the ratio between the vertical and horizontal dimensions of the box is 1.5. Note that this procedure can process faces of different sizes.

3.2 Cropping

The cropping stage refines the output coming from the location stage using DT. The DT are learned ('induced') from both positive ('face') and negative ('non-face') examples expressed in terms of extracted features such as entropy, mean, and standard deviation (sd). The detection stage is run across 8 x 8 windows from BOX1.

Each 8 x 8 window and its corresponding four quadrants is processed to yield the features required for learning from examples and the derivation of the DT. Entropy, mean, and standard deviation (sd) features are derived using original image data or Laplacian processed image data from the face box to yield thirty (30 = 5 x 3 x 2) features values.

Once features values become available learning from examples will induce DT. As inductive learning requires symbolic data, each one of the positive examples corresponding to the face region is tagged 'CORRECT', while the negative or counter-positive examples corresponding to non-face regions are tagged as 'INCORRECT'. Decision trees are induced using C4.5 [28]. The input to the C4.5 consists of a string of learning events, each event given as a vector of 30 attribute values. C4.5 takes a set of positive examples and a set of negative examples and builds a classifier as a decision tree whose structure consists of

- *leaves*, indicating class identity, or
- *decision nodes* that specify some test to be carried out on a single attribute value, with one branch for each possible outcome of the test.

A decision tree is now used to classify each 8 x 8 non - overlapping window from BOX1 by starting at the root of the tree and moving through it until a leaf is encountered. At each non leaf node a decision is made using an optimal (entropy) criterion such as the gain ratio criterion [Quinlan, 1986]. As a result of this procedure each 8 x 8 window from BOX1 is now labeled as face or non - face. The data set used to train the face comes from twelve (12) images whose resolution is 256 x 384 and consists of 2759 windows (CORRECT/ positive '**+**') examples, and 15673 (INCORRECT / negative '**-**') examples. Note that inducing the DT corresponds to training and requires suitable strategies. The (corrective) training strategy used here, that of tuning the DT on learning events the DT fails to resolve, corresponds to active learning [Krogh and Vedelsby, 1995]. The results reported (in Sect. 4) come from the application of (frozen) DT only on unseen test data.

3.3 Post Processing

The output from the cropping stage consists of labeled 8 x 8 (face vs non - face) windows. Horizontal and vertical profiles count the number of face labels and are thresholded to yield the extent of the face image, if present at all, and provide a new and refined approximation for the face as BOX2. Small counts in both the horizontal and vertical profiles indicate the absence of a face. Normalization takes BOX2, if found at all, and produces a standard size BOX3, available for further processing.

4. Experiments

Experiments were carried out to assess the feasibility of our novel algorithm for deciding if a face is present, and if it is present to crop it and normalize it. We briefly describe the data base used for our experiments and then proceed to report on face detection. For the purpose of our application we use the FERET facial data base [Gutta et. al., 1995] consisting of 1, 109 sets comprising 8, 525 images. Since large amounts of images were acquired during different photo sessions, the lighting conditions and the size of the facial images can vary. The diversity of the FERET data base is across gender, race, and age. The facial image sets were acquired without any restrictions imposed on expression and with two frontal images shot at different times during the photo session.

The data base for our experiments consists of 2,340 frontal face images coming from 817 subjects and 190 duplicate subjects. Each image is of size 256 x 384 using 256 gray scale levels. In order to assess the feasibility of our algorithm to reject images for whom no face is present we assembled a small data base consisting of 25 images of varied but complex backgrounds, whose size is 320 x 240 using again 256 gray scale levels. Examples of face and non - face images are shown in Fig. 2a and 2b, respectively.

The whole process for one face image is shown in Fig. 3. It display first the original face image, then it shows the intermediate outputs, BOX1, labeled contents of BOX1 and the corresponding horizontal and vertical profiles, the refined BOX2, and finally the normalized BOX3.

The performance obtained, stage wise, is as follows. Based on visual observation, the accuracy (head within the box) of the location (1st) stage is 99% - 22 incorrect boxes (out of 2340 image) were produced. The accuracy for the whole process is based on the visual observation that the box includes both eyes, nose, and mouth, and that the top side of the box is below the hairline. Note that incorrect as

(a) Face imges

(b) Non-Face images

Figure 2. Face and Non - Face Images

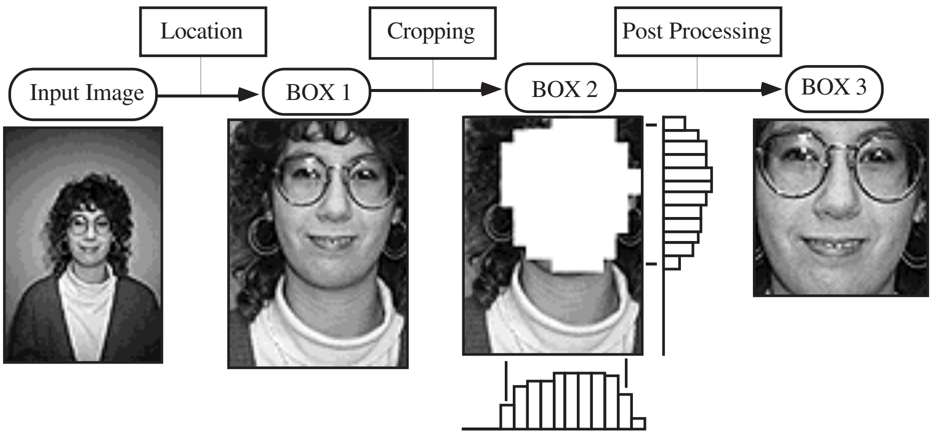

Figure 3. Face Detection Process

well as correct boxes are passed as BOX1 to the cropping stage. The overall accuracy at the end of the process is 96%. Fig. 4 shows the final result, BOX3, if at all, for the image examples shown in Fig. 2. The last image shown in Fig. 4a is one example for face mis detection and it corresponds to the last image from Fig. 2b.

5. Conclusions

This paper proposed a novel algorithm for face detection using decision trees (DT) and shows its generality and feasibility using a data base consisting of 2,340 face images from the FERET data base (corresponding to 817 subjects and including 190 sets of duplicates) over a semi uniform background.The algorithm decides first if a face is present, and if the face is present it crops ('box') the face. The algorithm used does not require multiple (scale) templates and the accuracy achieved is 96%. Accuracy regarding cropping the face is based on the visual observation that the box includes both eyes, nose, and mouth, and that the top side of the box is below the hairline. Experiments were also performed to assess the accuracy of the algorithm in rejecting images where no face is present. Using a small data base of 25 images of various but complex backgrounds the algorithm failed on two images for an overall accuracy rate of 92%.

(a)

(b)

Figure 4. Examples of (a) Face and (b) Non - Face Detection

6. References

Burel, G. and Carel, D. (1994), Detection and Localization of Faces on Digital Images, *Pattern Recognition Letters*, Vol. 15, 963-867

Chellappa, R., Wilson, C. L. and Sirohey, S. (1995), Human and Machine Recognition of Faces: A Survey, in *Proceedings of IEEE*, Vol. 83, No. 5, 705-740.

Gutta, S., Huang, J., Singh, D., Shah, I., Takacs, B. and Wechsler, H. (1995), Benchmark Studies on face Recognition, in *Proceedings of the International Workshop on Automatic Face and Gesture-Recognition (IWAFGR)*, M. Bischel (Ed.), 261-264, Zurich.

Hildreth, E. and D. Marr. (1980), Theory of Edge Detection, *Proceedings of Royal Society of London*, Vol. 207, 187-217.

Krogh, A. and Vedelsby, J. (1995), Neural Network Ensembles, Cross Validation and Active Learning, in *Advances in Neural Information Processing Systems (NIPS)*, D. S. Touretzky (Ed.), Vol. 2, 231-238, Morgan Kaufmann.

Quinlan, J.R. (1986), The Effect of Noise on Concept Learning, in *Machine Learning: an Artificial Intelligence Approach* 2, R.S. Michalski, J.G. Carbonell, and T.M. Mitchell (Eds.), 149-166, Morgan Kaufmann.

Rowley, H. A., Baluja, S. and Kanade, T. (1995), Human Face Detection in Visual Scenes, *Technical Report*, CMU-CS-95-158.

Samal, A. and Iyengar, P. (1992), Automatic Recognition and Analysis of Human faces and Facial Expressions: A Survey, *Pattern Recognition*, Vol. 25, No. 5, 65-77.

Yang, G. and Hunag, T. S. (1993), Human Face Detection in a Complex Background, *Pattern Recognition*, Vol. 27, No. 38, 53-63.

Acknowledgements: This work was partly supported by the US Army Research Lab under contract DAAL01-93-K-0099.