PARTS-BASED FACE DETECTION AT MULTIPLE VIEWS

Andreas Savakis and David Higgs
Department of Computer Engineering, Rochester Institute of Technology, Rochester, NY 14623, USA

Keywords: Face detection, parts-based, multiple views, neural network, Bayesian network.

Abstract: This paper presents a parts-based approach to face detection, that is intuitive, easy to implement and can be used in conjunction with other image understanding operations that use prominent facial features. Artificial neural networks are trained as view-specific parts detectors for the eyes, mouth and nose. Once these salient facial features are identified, results for each view are integrated through a Bayesian network in order to reach the final decision. System performance is comparable to other state-of the art face detection methods while providing support for different view angles and robustness to partial occlusions.

1 INTRODUCTION

Face detection in images and video is very important for many applications including surveillance, biometrics, content-based image retrieval and human-computer interaction. Image and video understanding systems often perform multiple tasks in addition to face detection, such as pose estimation, gaze detection, recognition of facial expressions, etc. Therefore, it would be helpful to develop face detection methods that utilize features which can be shared with other modules.

Face detection approaches may be categorized as feature-based or image-based (Hjelmas and Low, 2001). For example, neural networks and support vector machines were used in (Rowley et al., 1998) and (Osuna et al., 1997) without considering features. In feature based approaches, primitive parts can be utilized to match known object relationships, as in (Yow and Cipolla, 1997) and (Hsu et al., 2002). In (Yow and Cipolla, 1997), a Bayesian network was used to detect faces given evidence on the presence of features. Boosting of features was employed in (Viola and Jones, 2001) and (Xiao et al., 2003). The parts selection process can be automated using statistical methods, as in (Schneiderman and Kanade, 2004), (Weber et al., 2000), and (Fergus et al., 2003).

Automatic parts selection works without the need of expert knowledge to identify important features, but it does not always result in intuitively obvious regions. This makes it difficult to share detected features with other image understanding tasks.

Detection at multiple views may be accomplished as an extension of existing methods by training a single detector with all available views (Pontil and Verri, 1998), (Yang et al., 2000) or by training one model per viewpoint (Schneiderman and Kanade, 2000). A constellation model that is suitable for multipose detection was employed in (Weber et al. 2000), and (Fergus et al., 2003).

In this paper, we present a parts-based approach to face detection which utilizes prominent features that may be used by other image understanding methods. Thus, the results of feature detection modules can be shared between various tasks for efficient operation of the overall system. Furthermore, the parts-based approach allows detection under partial occlusion and is extended to incorporate multiple views.

2 FACE DETECTION SYSTEM

The basic framework for the parts-based face detection system is shown in Figure 1. This system contains separate modules, where each module is designed to support a different object view. Individual parts detectors are selected to correspond to prominent facial features, namely eyes, mouth and nose. Parts detectors for each view are separately trained and their outputs are fed into a Bayesian network arbitrator that determines whether a face is present at a given view.

Multi-layer artificial neural networks (ANNs) are used for parts detection. The input layer corresponds

Savakis A. and Higgs D. (2007).
PARTS-BASED FACE DETECTION AT MULTIPLE VIEWS.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IU/MTSV*, pages 298-301
Copyright © SciTePress

to a rectangular aperture at the selected reference scale, as shown in Figure 2. The ANN detector is applied as a sliding window, such that the network response at each location is recorded in an activation map. To accommodate moderate variations in scale, three window sizes are scaled to the ANN resolution.

The overall face detection system is shown in Figure 3. Histogram equalization is performed on each image subwindow to compensate for lighting variations. The equalized region is processed by the neural network parts detectors and an activation map is obtained for each view. The activation map is lowpass filtered to reduce the effects of outliers, and the facial feature locations are identified based on the maximum values of the activation map.

The results of the feature detector are interpreted by the arbitrator network. Bayesian networks were chosen as view arbitrators because of their natural resistance to overfitting and ability to incorporate incomplete data and cause-and-effect relationships (Heckerman, 1995). This required discretizing the output of the neural networks to indicate presence or absence of a feature.

The final decision incorporates the detection results at different poses and this is done with a simple logical OR operation, since little additional leverage can be gained by applying a more complex decision-making process.

3 RESULTS

The FERET database (Phillips, 2000) was used to construct training and testing sets, as it contains multiple viewing angles of human faces. Other desirable characteristics include a large number of subjects and good diversity across age, race, and gender. There were 1364 images of the Frontal A view and 690 images of the Quarter Left view that selected to illustrate the capabilities of the system, henceforth referred to as “frontal” and “side” views.

The four facial features selected were the left eye, right eye, nose, and mouth. Both the frontal and side views allow for all four facial features to be visible. The input images were scaled so that most parts could be represented with 400 or less input neurons. Eyes were detected based on a 12x20 pixel aperture, with the exception of the right eye of the side view; it was 12x16 pixels due to foreshortening effects. The nose and mouth detectors were 18x20 and 14x32 pixel windows, respectively. The resulting faces had approximate dimensions of 50 pixels high by 55 pixels wide.

The bootstrapping process was used for training the parts detectors (Sung and Poggio 1998). The component parts were extracted from the training images, preprocessed, and added to the training set as positive examples. To prevent an initial bias, an equally sized negative set was added by taking random preprocessed subwindows from a background dataset consisting of 451 images from the Caltech background image database.

Arbitration networks for each view were trained in order to make a decision about the presence or absence of a face in the scene. Part activation values were gathered and used to determine event thresholds for each part. The threshold was selected by plotting a receiver operating characteristic curve (ROC curve) for the activation values and choosing the optimal threshold point.

The use of Bayesian networks allowed for the inaccuracies of the individual neural networks to be handled implicitly by the network itself. The conditional probability tables (CPTs) were determined experimentally from the training set. Since the true presence of an object was known prior to finding the part detections, statistics were gathered for each part in the form of CPTs. Table 1 shows how the results of training the individual parts detectors were incorporated in CPTs. The correspondence between each part and its associated view was found by counting the frequency of detection with respect to whether the candidate image contained a face or not. The experimental part detection rates conformed to expectations, in that parts are typically detected when a view was present and not detected when the view is not present. Bayes’ rule was applied to each view’s arbitration network to find an expression relating the presence of an object at a certain view to the conditional part probabilities, as shown in the equation below:

$$P(v \mid d_1, d_2, d_3, d_4) = P(v)P(d_1 \mid v)P(d_2 \mid v)P(d_3 \mid v)P(d_4 \mid v)$$

where v represents a view and d’s represent parts detected. For a given set of part detections, the equation was evaluated twice – once for each state of the view detection – substituting in the corresponding CPTs for each part. The view state with the larger network probability was the view belief for the image. n most cases, two or more of the four parts at any particular view indicated the presence of the object.

The validation process was to simply run the standard detection scheme on the validation images of each cross-validation set. The testing results illustrate the performance of the system on images that it had never encountered. The output of the final stage was a binary decision between face and non-face; the average performance is shown in Table 2. The overall detection performance is better than the performance of any of the individual part detectors, which demonstrates the strength of Bayesian decisions in this context. Side face detection performed slightly better on average than the frontal face detection, which could be expected by comparing the part CPTs of each view.

For demonstration purposes, the proposed parts-based face detection method was applied to subjects outside the FERET database. Figure 4 shows two correctly detected faces that are at different scales and varying lighting conditions. Note that occlusion of one eye did not affect the detection result.

4 CONCLUSIONS

This paper presents a parts-based face detection approach that includes support for multiple viewing angles. Parts detectors for eyes, mouth and nose were implemented using neural networks trained using the bootstrapping method. Bayesian networks were used to integrate part detections in a flexible manner, and were trained on a separate dataset so that the experimental performance of each part detector could be incorporated into the final decision.

Images from the FERET human face database were selected for training and testing. Individual part detection rates ranged from 85% to 95% against testing images (Table 1). Cross-validation was used to test the system as a whole, giving average view detection rates of 96.7% and 97.2% respectively for the frontal and side views, and an overall face detection rate of 96.9% (Table 2). A 5.7% false-positive rate was demonstrated on background clutter images.

Table 3 shows that the approach presented in this paper performs in a manner comparable to other research efforts within the field of face detection, with minimal restrictions that would hinder generalization to other object categories. In addition, this approach provides the additional benefit of support for different view angles. Finally, selecting prominent facial features for face detection provides a benefit for other image understanding modules that may utilize the detected features.

ACKNOWLEDGEMENTS

This research was sponsored in part by the Eastman Kodak Company and the Center for Electronic Imaging Systems (CEIS), a NYSTAR-designated Center for Advanced Technology in New York State.

REFERENCES

Fergus R., Perona P., and Zisserman A., 2003. Object Class Recognition by Unsupervised Scale-Invariant Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Heckerman D., 1995. A Tutorial on Learning With Bayesian Networks. Technical Report MSR-TR-95-06, Microsoft Research, Advanced Technology Division.

Hjelmas, E. and Low, B. K., 2001. Face Detection: A Survey. In *Computer Vision and Image Understanding*, vol. 83, pp. 236–274.

Hsu R. L., Abdel-Mottaleb M., and Jain A. K., 2002. Face Detection in Color Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 696–706..

Osuna E., Freund R., and Girosi F., 1997. Training Support Vector Machines: an Application to Face Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 130–136.

Phillips P. J., Moon H., Rizvi S. A., and Rauss P. J., 2000. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 090–1104, October.

Pontil M. and Verri A., 1998. Support Vector Machines for 3D Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 637–646.

Rowley H. A., Baluja S., and Kanade T., 1998. Neural Network-Based Face Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 23–38.

Schneiderman, H. and Kanade T., 2000. A Statistical Method for 3D Object Detection Applied to Faces and Cars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1746–1759.

Schneiderman H. and Kanade T., 2004. Object Detection Using the Statistics of Parts. *International Journal of Computer Vision*, vol. 56, pp. 151–177.

Sung K. K. and Poggio T., 1998. Example-based Learning for View-based Human Face Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 39–51.

Viola P. A. and Jones M. J., 2001. Rapid Object Detection using a Boosted Cascade of Simple

Features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 511–518.

Weber M., Welling M., and Perona P., 2000. Unsupervised Learning of Models for Recognition. In *Proceedings of the European Conference on Computer Vision*, vol. 1, pp. 18–32.

Xiao R., Zhu L., and Zhang H. J., 2003. Boosting Cascade Learning for Object Detection. In *IEEE International Conference on Computer Vision, ICCV03*.

Yang M. H., Roth D., and Ahuja N., 2000. Learning to Recognize 3D Objects with SNoW. In *Proceedings of the European Conference on Computer Vision*, vol. 1, pp. 439–454.

Yow, K. C. and Cipolla R., 1997. Feature-based human face detection. *Image and Vision Computing*, vol. 15, pp. 713–735.

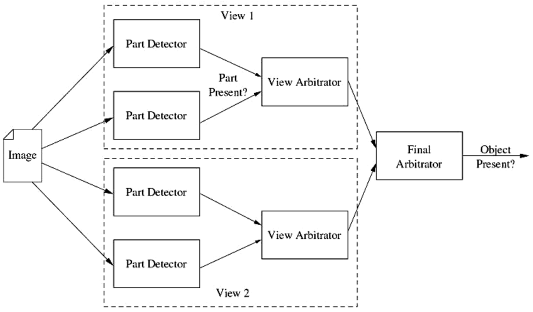

Figure 1: Framework for parts-based face detection at multiple views.

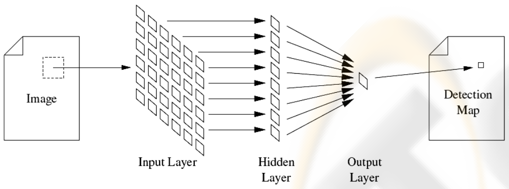

Figure 2: Neural network parts detector.

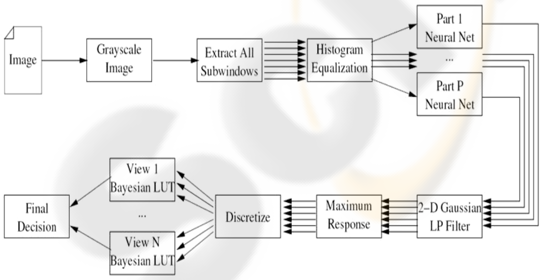

Figure 3: Overall Flow of Face Detection System.

Table 1: Conditional Probability Tables (CPTs) based on the performance of feature detectors.

Frontal View State	P(Left Eye = T)	P(Right Eye = T)	P(Nose = T)	P(Mouth = T)
F	0.0798	0.1079	0.0458	0.0790
T	0.8964	0.8624	0.9519	0.8663

Side View State	P(Left Eye = T)	P(Right Eye = T)	P(Nose = T)	P(Mouth = T)
F	0.0791	0.0924	0.0658	0.1145
T	0.9111	0.8560	0.9227	0.8449

Table 2: Face Detection Performance.

	Detected Images		Detected Percent	
Dataset	Non-Face	Face	Non-Face	Face
Frontal	11.1	329.9	3.3%	96.7%
Side	4.9	167.6	2.8%	97.2%
Background	106.3	6.4	94.3%	5.7%
Face Overall	**16**	**497.5**	**3.1%**	**96.9%**

Table 3: Comparison with existing work.

	Model	Detector(s)	Arbitrator(s)	Detection
Sung & Poggio (1998)	Image	PCA Distance Metric	Neural Network	96.7%
Osuna *et al.* (1997)	Image	SVM	Threshold	97.1%
Rowley *et al.* (1998)	Image	Neural Networks	Threshold	99.5%
Fergus *et al.* (2003)	Parts	Gaussian Distribution	Statistical Model	96.4%
Yow & Cipolla (1997)	Parts	Edge Detection	Bayesian	92.9%
View Based Parts	**Parts**	**Neural Networks**	**Bayesian**	**96.9%**

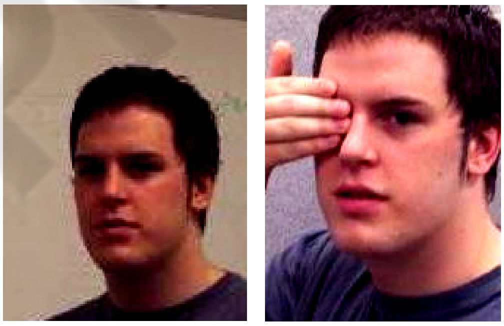
(a) (b)

Figure 4: Examples of detected faces showing robustness to scale variations, lighting variations and occlusion.