Face Detection Using Hierarchical Integral Projections

Ming Li, Baozong Yuan, Ming Liu

Institute of Information Science, Beijing Jiaotong University, Beijing, China 100044

Abstract

This paper describes a novel density-based clustering method for multiple frontal face detection in grey-scale images. This method models the distribution of the face and non-face pattern by means of a few density-based clusters. To detect faces in an input image, the algorithm matches window patterns at different image location and scales against the distributed face and non-face model patterns. This method showed good performance in the experiments.

Keyword: Face detection, clustering, integral projection

1. Introduction

Face detection and recognition are the most challenging tasks for visual form analysis and object recognition; for detailed surveys see [1], [2] and [3].A successful automatic face recognition system should have powerful face detection system. Face detection's task is to determine the location of human faces in an input image. It generally learns the statistical models of the face and non-face images, and then applies a two-class classification rule to discriminate between face and non-face patterns [4].

Lin et al. presented a face detection system using probabilistic decision-based neural network (PDBNN) [5]. Their method firstly extracts feature vectors based on intensity and edge information of the facial region that contains eyebrows, eyes and nose. Then, the obtained feature vectors are fed into two PDBNNs and the fusion of the output determined the classification result. Rowley et al use a multilayer neural network to learn the face and non-face patterns form face/non-face images [6], and their work achieved significant performance. Despite the Neural Network's success in face detection, one drawback still exist: the network architecture has to be extensively tuned (number of layers, number of nodes, learning rates, etc.) to get exceptional performance.

Sung and Poggio developed a distribution-based system for face detection [7]. They models the distribution of human face pattern by means of six face and six non-face model clusters. Yang's method use a mixture of use Factor analysis (FA) for modeling the covariance structure of training images [8]. Liu [4] combine the input image, its 1D Harr wavelet representation and its amplitude projections to form a discriminating feature vector. Then, he use multivariate normal distribution to model the face and non-face patterns and obtained a Bayesian Classifier. Osuna et al. use Support Vector Machine (SVM) to detect face in complex scene and achieve good result [10]

This paper presents a hierarchical integral projection-based clustering method for multiple frontal face detection by integrating feature analysis, modeling, and classification. The main contributions of the paper come from 1) the hierarchical projection feature analysis of the input images, 2) cluster analysis of face and non-face patterns.

Our method is trained using the image patches, which cropped from FERET face images [9] and ORL face database (http://www.cam-orl.co.uk) and several natural images. The training set contains 6737 images (containing a total of 2989 face patches), which are normalized to a standard resolution, 19×19 . To enlarge our training set, we also added the mirror images of the face images into training set.

The organization of this paper is as follows: In Section 2, we propose the Hierarchical Integral Projections vector for face detection. In Section 3, we analyze the face and non-face patterns with Vector Quantization (VQ) algorithm.

We compare 2D-LDA with eigenfaces, fisherfaces and 2DPCA on the ORL face database. Finally, the paper concludes with some discussions in Section 4.

2. Hierarchical Integral Projections

Integral projections are important image feature, can be used to represent the visual appearance of a certain kind of object under a relatively wide range of conditions, i.e., to model object classes. In this way, object analysis can be done by fitting a test sample to the projection model. In face detection, integral projection are able to capture the vertical symmetric

distributions an the horizontal characteristics of human face images.

Firstly, we give out the definitions on integral projections.

Let $i(x,y) \in \Re^{m\times n}$ be a greyscale image and $R(i)$ a region in this image, i.e., a set of contiguous pixels in the domain of i. The horizontal and vertical integral projections of $R(i)$, denoted by $P_{HR}(i)$ and $P_{VR}(i)$ respectively, are discrete and finite 1-D signals give by

$$P_{HR(i)} : \{x_{min},\ldots,x_{max}\} \to \mathbf{R}; \quad (1)$$

where $P_{HR(i)}(x) = |Rx(i)|^{-1} \sum_{y\in Rx(i)} i(x,y)$.

$$P_{VR(i)} : \{y_{min},\ldots,y_{max}\} \to \mathbf{R}; \quad (2)$$

where $P_{VR(i)}(y) = |Ry(i)|^{-1} \sum_{y\in Ry(i)} i(x,y)$.

$$x_{min} = \min_{(x,y)\in R(i)} x; x_{max} = \max_{(x,y)\in R(i)} x;$$
$$y_{min} = \min_{(x,y)\in R(i)} y; y_{max} = \max_{(x,y)\in R(i)} y; \quad (3)$$

$$Rx(i) = \{y | \forall y, (x,y) \in R(i)\};$$
$$Ry(i) = \{x | \forall x, (x,y) \in R(i)\}. \quad (4)$$

The sets $\{x_{min},\ldots,x_{max}\}$ and $\{y_{min},\ldots,y_{max}\}$ are called the domains of the horizontal and vertical integral projection, denoted by $Domain(P_{HR(i)})$ and $Domain(P_{VR(i)})$ respectively.

2.1. Hierarchical Integral Projections of Face Images

To analyze face images, we have some preprocess to do in advance. We apply an oval mask to eliminate some near-boundary pixels of each window pattern. Because these masked pixels are mostly background pixels, removing them ensures that we do not wrongly introduce any unwanted background information.

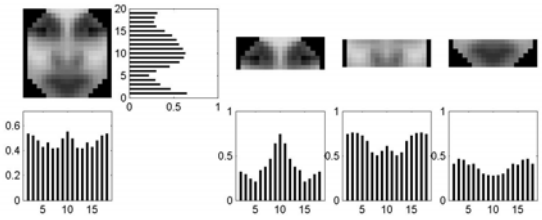
According to the definitions of integral projection, the horizontal (row) and vertical (column) projections of $i(x,y)$ can be calculated, denoted by vectors P_{HA} and P_{VA} respectively. Firstly, we apply the oval masks to the images, and the value of mask pixel is 0. Let $I(x,y)$ be the masked version of $i(x,y)$, we have

$$P_{HA}(y) = |Ry|^{-1} \sum_{x=1}^{19} I(x,y), \quad 1 \le y \le 19, \quad (5)$$

$$P_{VA}(x) = |Rx|^{-1} \sum_{y=1}^{19} I(x,y), \quad 2 \le x \le 18. \quad (6)$$

The mean face of face pattern, its horizontal and vertical integral projections are showed in Fig.1. (a).

To describe the face patterns more precisely, the face images are divided into 3 regions in vertical orientation: eyes & eyebrows region, nose & cheeks region and mouth region. Then, the vertical projections of these three sub-regions, P_{VE}, P_{VN} and P_{VM} are obtained. The three sub-regions and their vertical projections are shown in Fig.1. (b).



(a) (b) (c) (d)

Fig.1. Hierarchical integral projections of the mean face. Integral Projections

The integral projections only describe the intensity property of the face images. That is not enough for face recognition. The 1D Harr representation of $I(x,y)$ is defined as $I_H(x,y) \in \Re^{(m-1)\times(n-1)}$,

$$I_H(x,y) = \left(|I(x+1,y) - I(x,y)| + |I(x+1,y+1) - I(x,y+1)|\right) + \left(|I(x,y+1) - I(x,y)| + |I(x+1,y+1) - I(x+1,y)|\right)$$
$$1 \le x \le 18,\ 1 \le y \le 18 \quad (7)$$

After study on the mean wavelet-face, we can find that the eye, nose and mouth regions have large amplitude. We divide the wavelet-face into 3×3 sub-regions. So, we can defined the 2D integral projections for images,

$$P_{R(m,n)} = \frac{1}{Area(R(m,n))} \sum_{y\in R(m,n)} \sum_{x\in R(m,n)} I_H(x,y), \quad (8)$$
$$1 \le m \le 3, \quad 1 \le n \le 3$$

So, we can obtain a Hierarchical Integral Projections vector $P \in \Re^N$ by concatenation of the vectors P_{HA}, P_{VA}, P_{VE}, P_{VN}, P_{VM} and $P^t_{R(m,n)}$.

$$P = (P^t_{HA}, P^t_{VA}, P^t_{VE}, P^t_{VN}, P^t_{VM}, P^t_{R(m,n)})^t \quad (9)$$

where t is the transpose operator.

3. Modeling of "Face" Patterns

Given an image, to discriminate whether it is a face pattern is rather difficult. This is because of the wide range of permissible pattern variations in face images. For example, for 19×19 greyscale images, there are $256^{19\times19}$ possibilities. It is impossible to describe the face pattern exactly in such a high-dimensional space by enumerating all the face images. Instead, we can use the statistical method to obtain a coarse but fairly reliable representation of the actual face manifold from a rather large image database.

To model the distribution of "face" pattern precisely, a Batch Learning Vector Quantization (BLVQ) method is proposed. We introduce the idea of Vector Quantization (VQ) firstly.

3.1. Vector Quantization (VQ) algorithms

VQ is proposed as a method of lossy compression that applies statistical techniques explicitly to optimize distortion/bit rate trade-offs [11]. For each k-dimensional input vector $\mathbf{X}$, the VQ encoder α determines the best match from a collection of N reproduction vectors or codewords, $C = \{\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \ldots, \hat{\mathbf{X}}_N\}$ and output the binary representation of the chosen codeword's index: $\alpha(\mathbf{X}) = i$ if $\hat{\mathbf{X}}_i$ is selected. VQ decoder β reverses this process as it has the index i as input and outputs the reproduction $\hat{\mathbf{X}}_i = \beta(i)$. Then, VQ is described by the encoder-decoder combination (α, β).

Given the compressed dimension k, we can obtained the optimal VQ resolution by minimize the mean squared error (MSE). For a set of sample vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_L\}$, the average distortion measure is

$$D = \frac{1}{L}\sum_{i=1}^{L} d(\mathbf{x}_i, \beta(\alpha(\mathbf{x}_i))), \qquad (10)$$

where $d(\mathbf{x}, \mathbf{y})$ represents the Euclidean distant between **x** and **y**.

VQ is an unsupervised clustering process. If it is applied to classification, the class labels can be assigned to each codeword and NN-like classifiers can be employed. Kohonen [12] proposed a supervised VQ algorithm for classification: Learning Vector Quantization (LVQ). LVQ assigns each codeword with a class label according to the distribution of classes and ensure that the estimated class borders agree with Bayes Borders. LVQ is an on-line algorithm that has the following update equation:

$$\mathbf{m}_i[n+1] = \begin{cases} \mathbf{m}_i[n] + \alpha[n](\mathbf{x}[n+1] - \mathbf{m}_i[n]) & \text{if } \mathbf{x}[n+1] \in \mathbf{R}_i[n] \text{ and } \mathbf{x}[n+1], \mathbf{m}_i \in \text{sameclass} \\ \mathbf{m}_i[n] - \alpha[n](\mathbf{x}[n+1] - \mathbf{m}_i[n]) & \text{if } \mathbf{x}[n+1] \in \mathbf{R}_i[n] \text{ and } \mathbf{x}[n+1], \mathbf{m}_i \notin \text{sameclass} \end{cases} \quad i = 1, \ldots, K \qquad (11)$$

where $\alpha[n]$ is the step size function that is limited to $(0,1)$, $\mathbf{m}_i[n]$ is the *i*-th codeword at time n.

3.2. Batch Learning Vector Quantization (BLVQ)

Kohonen's LVQ algorithm are statistical learning procedures that tend to perform VQ over a probability density function that has zero points at Bayes borders. To speed up the training process, we propose a batch version LVQ algorithm.

Table 1 is the summary of the training process of BLVQ.

- **Input**: $(\mathbf{x}_k, y_k) \in \mathbf{T}$ ($k = 1, \ldots, N$; $y_k = 1$, face images; $y_k = -1$, non-face), step size parameter α $(0 < \alpha < 1)$ and δ (the threshold of movement).
- **Initialization**: *m* face images and *n* non-face images selected randomly as face codeword $C'_{1,i}$ $(i = 1, \ldots, m)$ and non-face codeword $C''_{1,j}$ $(j = 1, \ldots, n)$.
- At *t*-th iteration:
  1. Calculate the Euclidean distances between training data and codewords, $D'_{k,i} = \left\|\mathbf{x}_k - C'_{t,i}\right\|^2$ and $D''_{k,j} = \left\|\mathbf{x}_k - C''_{t,j}\right\|^2$.
  2. Form a (nearest-neighbour-based) partition $\mathbf{P}_{m+n} = \bigcup_{l=1}^{m+n} \mathbf{R}_l$ by assigning each training data to the nearest codeword.
  3. Update each face codeword,
  $$C'_{t+1,i} = C'_{t,i} + \alpha^t \cdot \frac{1}{N_i} \sum_{\mathbf{x}_k \in \mathbf{R}_i} y_k \cdot (\mathbf{x}_k - C'_{t,i}) \; (i = 1, \ldots, m)$$
  4. Update each non-face codeword,
  $$C''_{t+1,j} = C''_{t,j} + \alpha^t \cdot \frac{1}{N_j} \sum_{\mathbf{x}_k \in \mathbf{R}_j} (-y_k) \cdot (\mathbf{x}_k - C''_{t,j}) \; (j = 1, \ldots, n)$$
  5. Calculate the movement of seeds,
  $$movement = \sum_{i=1}^{m} \left\|C'_{t+1,i} - C'_{t,i}\right\|^2 + \sum_{j=1}^{n} \left\|C''_{t+1,j} - C''_{t,j}\right\|^2.$$
  6. If $movement \geq \delta$, increase *t* by 1 and go to Step 1.
- **Output:** the codewords of face patterns and non-face patterns.

Table 1. The pseudocode of BLVQ algorithm.

Consider the non-face images distribute more dispersively than face images in the feature space, we usually choose $m < n$.

By using BLVQ, we can got a $m+n$ codebook to encode the face patterns and non-face patterns. Then, we can obtain a k-Nearest Neighbor (kNN) classifier which caters to Bayes border.

4. Experimental Results

Our algorithm can detect upright frontal views of human faces with little rotation to the left and to the right. It searches the input images for faces of small scales first, then searches for larger scales in different scale versions of the original images. Using this strategy, our system can detect human faces in different scales.

Each time a face pattern is found, our algorithm draws an appropriately sized box at the corresponding location in the output image. Because the same face

Table 2
Classification Correct Rates on the CBCL face database #1

Size of Codebook	Classification Correct Rates	
	Face	Non-Face
6 Face & 10 Non-Face	82.97%	96.29%
10 Face & 15 Non-Face	89.66%	98.80%
15 Face & 20 Non-Face	90.31%	99.01%

pattern can be detected at multiple scales and at slightly shifted location, there can be multiple boxes enclosing each face pattern in the output image. An arbitration strategy is applied to get rid of the overlapped boxes.

We firstly applied our algorithm on the CBCL Face Database #1 (http://www.ai.mit.edu/projects/cbcl), provided by MIT Center For Biological and Computation Learning. This set contains 2,429+472 faces images and 4,548+23,573 non-faces. All these images are cropped to 19×19 size. Table 2 shows the detection rates with different codebook size.

5. Conclusions

This paper presents a novel method, which the face image's integral projections as the key features for face detection. We combine the hierarchical integral projections and 1D-Harr wavelet to present the face images. We also propose a batch version LVQ algorithm, BLVQ, to model the face and non-face patterns. The experimental results show that our algorithm is efficient. Our system achieves 92.04% correct detection rate and 6 false detections.

But there still exist some problems to be resolved in future. For example, the algorithm scans the input images iteratively in different scale and position. This makes the computation-cost is rather high. If we can filter out the background patches which dislike face very much, the algorithm's can be promoted further.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 60441002) and the University Key Research Project (No. 2003SZ002).

6. References

[1] M.Yang, D.J.Kriegman and N.Ahuja, "Detecting Faces in Images: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, Jan. 2002, 34-58.

[2] E. Hjelmas and B.K.Low, "Face Detection: A Survey," *Computer Vision and Image Understanding*, vol. 83, pp. 236-274, 2001.

[3] R. Chellappa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol.83, no. 5, pp 705-740, 1995.

[4] C.Liu, "A Bayesian Discriminant Features Method for Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, June. 2003, 725-740.

[5] S.H.Lin, S.Y.Kung and L.J.Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, Jan. 1997.

[6] H. Rowley, S. Baluja and T. Kanade, "Neural Network-Based Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-38, Jan. 1998.

[7] K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39-51, Jan. 1998.

[8] M. Yang, N. Ahuja and D. Kriegman, "Mixtures of Linear Subspaces for Face Detection," *Proc. Fourth Int'l Conf. Automatic Face and Gesture Recognition*, pp. 70-76, 2000.

[9] P. J. Philips, and H. Moon, S. A. Rizvi, and P. Rauss, "The FERET Testing Protocol," in *Face Recognition: From Theory to Applications* (H. Wechsler, P. J. Philips, V. Bruce, F. F. Soulie and T. S. Huang, eds.), Berlin: Springer-Verlag, pp. 244-261, 1998.

[10] E. Osuna, R. Freund and F. Girosi, " Training Support Vector Machines: An Application to Face Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 193-199, 1997.

[11] H.Abut, Ed., Vector Quantization, *IEEE Reprint Collection*. Piscataway, N.J.: IEEE Press, May 1990.

[12] T.Kohonen, *Self-Organization Maps*, Berlin: Springer-Verlag, 2nd ed., 1996.

[13] P.J. Philips, H. Wechsler, J. Huang, and P. Rauss, "The FERET Database and Evaluation Procedure for Face-Recognition Algorithms," *Image and Vision Computing*, vol. 16, pp. 295-306, 1998.