Face Detection and Synthesis Using Markov Random Field Models

Abstract

Markov Random Fields (MRFs) are proposed as viable stochastic models for the spatial distribution of gray level intensities for images of human faces. These models are trained using data bases of face and non-face images. The MRF models are then used for detecting human faces in test images. The number of human face images in the training data base can be increased by simulating face-like as well as non-face like images from the trained MRFs. These simulated images are added to the existing training data bases and the corresponding MRF parameters are re-estimated. We show that the resulting face detection algorithm detects a significantly less number of false positives. We investigate the performance of the face detection algorithm for two classes of MRFs given by the first and second order neighborhood systems.

Key words and phrases: Markov Random Fields, face detection, maximum pseudolikelihood estimation, simulated annealing, site permutation.

1. Introduction

Numerous attempts have been made in recent years to detect human faces in images using a variety of techniques . Some of the reported work in the literature include face detection algorithm based on neural networks [5, 6], tree classifiers [1], distance from prototype criteria [7] and Markov Chains [4, 2]. However, the inherent spatial nature of digital images makes MRFs a natural choice for modeling the distribution of gray level intensities on these images. Using MRF models for modeling and subsequent face detection has been reported in [3]. The results given in [3] indicate that the approach to detecting faces using MRFs look promising. However, due to computational complexity, the MRFs used in [3] were only a valid approximation. In this paper, we avoid approximations made in the previous paper; exact estimation procedures lead to parameter estimates for the actual MRF model. An immediate advantage is that now we are able to simulate face-like images from the estimated MRF models and increase the number of training samples in both the face and non-face data bases. Subsequent re-estimation of model parameters and detection show that significant reduction of error rates are obtainable for test images. We believe that our approach is novel in the sense that the proposed MRF models can be used for both detection and synthesis of face images.

The MRF models used here do not utilize high level feature extraction for the purpose of face detection. Indeed, our aim here is to provide an initial low-level detection algorithm. In the post processing stage, algorithms based on facial features can be utilized to finally decide if a face is indeed present in the test image. For this reason, we put greater emphasis in developing algorithms with low false negative rates in the detection framework. We also adopt the best discriminating MRF approach reported in [3]. Thus, the most discriminating permutation was found using the chi-square criteria proposed in [3], where error rates were shown to be significantly lower for the most discriminating permutation compared to the natural order of sites. See [3] for details.

The remainder of this paper is organized as follows. In Section 2, we present the MRF models for face detection. In Section 3, we estimate the MRF parameters based on maximizing the logarithm of the observed pseudolikelihood. Simulation of face like images and the performance of the face detection algorithm on real images is illustrated in Section 4.

2. The Markov Random Field Models

Let $S = \{1, 2, \cdots, \#S\}$ denote the collection of all sites in a $R \times C$ image, where $\#S = RC$. For each site s_0 in S, we denote by x_{s_0} to be the gray level intensity at that site (this is an integer between 0 and $L-1$, both inclusive, and where L is the number of gray levels). Also, we will denote by x_{-s_0} to be the gray level intensities of all sites in S excluding site s_0. The spatial distribution of gray level intensities, $\mathcal{X} = \{x_s, s \in S\}$ on S will be modeled as a Markov Random Field (MRF) with an associated neighborhood system $\mathcal{N} = \{N_s, s \in S\}$, where N_s denotes the neighbors of site s. We consider the first and second order neighborhood structure for the MRF models (see Figure 1). Thus, in Figure 1, $x_{s_0,2}, x_{s_0,4}, x_{s_0,7}$ and $x_{s_0,5}$ represent the north, west,

$x_{s_0,1}$	$x_{s_0,2}$	$x_{s_0,3}$
$x_{s_0,4}$	x_{s_0}	$x_{s_0,5}$
$x_{s_0,6}$	$x_{s_0,7}$	$x_{s_0,8}$

Figure 1. First and second order neighborhood systems for the site s_0 and corresponding gray level intensities.

south and east neighbors of x_{s_0}, respectively, for the first order neighborhood system. For the second order neighborhood system, the additional sites $x_{s_0,1}, x_{s_0,3}, x_{s_0,6}$ and $x_{s_0,8}$ are also defined to be neighbors of s_0.

The joint distribution of MRFs is uniquely defined by specifying of the conditional distribution (local characteristics) of x_{s_0} given its neighbors, $\{x_t, t \in N_{s_0}\}$, at each site s_0 in S. We consider the local charateristics at site s_0 given by

$$p(\,x_{s_0}\,|\,x_{-s_0}\,) = \frac{\exp\{H(x_{s_0}\,|\,x_{-s_0})\}}{\sum_{x_{s_0}=0}^{L-1} \exp\{H(x_{s_0}\,|\,x_{-s_0})\}}, \qquad (1)$$

where $H(x_s\,|\,x_{-s}) = \alpha_s x_s + \sum_{t\in N_s} \beta_{s,t} x_s x_t$ with parameters $\{\alpha_s, s \in S\}$ and $\{\beta_{st}, s \in S, t \in N_s\}$ for all sites s in S with neighbors $t \in N_s$. The joint distribution (likelihood) on S (provided $\beta_{st} = \beta_{ts}$) is given by

$$p(\underline{x}) = \frac{\exp\left\{\sum_s \alpha_s x_s + \sum_{s\sim t} \beta_{st} x_s x_t\right\}}{\sum_{x_1}\sum_{x_2}\cdots\sum_{x_{\#S}} \exp\left\{\sum_s \alpha_s x_s + \sum_{s\sim t} \beta_{st} x_s\, x_t\right\}}, \qquad (2)$$

where $s \sim t$ stands for all pairs of sites s and t that are neighbors in S. The normalizing constant in (2) is hard to handle computationally; thus, we resort to simpler approximations of the likelihood to avoid the normalizing constant. One such approximation is the pseudolikelihood (PL) defined by

$$\begin{aligned} PL &= \prod_{s=1}^{\#S} p(x_s\,|\,x_{-s}) \\ &= \prod_{s=1}^{\#S} \frac{\exp\left\{\alpha_s x_s + \sum_{t\in N_s} \beta_{st}\, x_s x_t\right\}}{\sum_{x_s=0}^{L-1} \exp\left\{\alpha_s x_s + \sum_{t\in N_s} \beta_{st} x_s x_t\right\}}. \qquad (3) \end{aligned}$$

3. Model Training

In [3], it was shown that the heterogeneous MRF model (corresponding to choosing different α_s and β_{st} for the different sites s and t in S leads to better detection results. Thus, we report the procedure to obtain parameter estimates based on training samples for the heterogeneous model. Also, the MRF models in Section 2 are fit to a permutation π_{opt} of sites in the image, π_{opt} being the most discriminating MRF model found using the chi-square criteria in [3].

The MRF models in Section 2 are trained using a database of faces and nonfaces. Face examples are generated by extracting gray level values from a 20×15 window (which contains the central part of the human face). Each pixel can take the 16 ($L = 16$) possible values of gray levels. The nonface examples are generated from images that resemble a face but are not actually so. The models were trained using 2000 sample images from each of the face and non-face training data base. Figures 2 and 3 each give 6 examples of face and nonface images in the training database.

For each training database (face and non-face), let N denote the total number of sample images in the data base. Thus, $N = 2000$ here. The value of the observed log pseudolikelihood (LPL) (see (3)) for the training sample is given as

$$\begin{aligned} &LPL(\alpha,\beta) = \\ &\quad \sum_{k=1}^{N}\sum_{s=1}^{\#S}\Bigg(\alpha_s x_s^{(k)} + \sum_{t\in N_s} \beta_{st}\, x_s^{(k)} x_t^{(k)} - \\ &\quad \log\Bigg(\sum_{x_s=0}^{L-1} \exp\{\alpha_s x_s + \sum_{t\in N_s} \beta_{st} x_s x_t^{(k)}\}\Bigg)\Bigg), \quad (4) \end{aligned}$$

where $\alpha = \{\alpha_s, s \in S\}$, $\beta = \{\beta_{st}, s \in S, t \in N_s\}$, and $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_{\#S}^{(k)})$ are the observed gray level intensities of the k-th training image. We seek the maximum pseudolikelihood (MPL) estimates, $\hat{\alpha}$ and $\hat{\beta}$, that maximize the observed LPL with respect to α and β. We assume toroidal (periodic) boundary conditions for the MRF defined for the sites permuted by π_{opt}. Thus, for the first order neighborhood system, there are a total of 900 unknown parameters; a total of 300 for the α_s, 300 for β_{st} along the horizontal direction and 300 for β_{st} along the vertical direction. For the second order neighborhood system, the number of unknown parameters is 1500 (300 for the α_s, and 1200 for the β_{st} for the 4 different directions in space). A multidimensional version of the Newton-Raphson iteration procedure is used to find the MPL estimates of α and β. One advantage of using the criteria in (4) is that resulting Hessian matrix for the iterative procedure is sparse; significant reduction in computational time and memory is thus achieved. The maximization algorithm was coded in MATLAB and was run on a PC with 256 MB of memory and processing speed of 733 MHz. The iterative procedure takes about 5 minutes to converge with 2000 training samples.

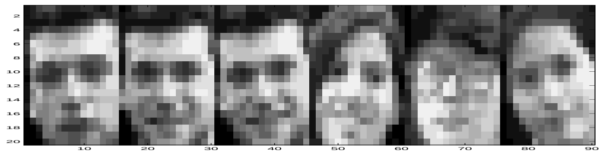


Figure 2. Examples of faces in the training data (20 × 15 images with 16 gray levels).

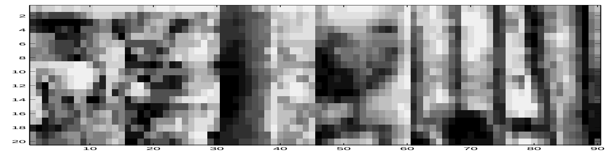
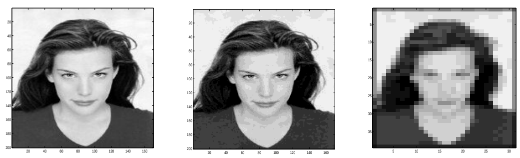


Figure 3. Examples of nonfaces in the training data (20 × 15 images with 16 gray levels).



Original image | Modulo 16 reduction | Scaled Image

Figure 4. Effects of Blocking and Scaling.

4. Face Detection and Synthesis

4.1. Face Detection Algorithm

A test image, $\{x_s^{inp}, s \in S\}$, is classified as a face if the log pseudolikelihood ratio (LPR) of face to non-face,

$$LPR = \sum_{s=1}^{\#S} \log \left(\frac{\hat{p}_{face}(x_s^{inp} \mid x_{-s}^{inp})}{\hat{p}_{nonface}(x_s^{inp} \mid x_{\pi^{-s}}^{inp})} \right) > 0. \quad (5)$$

Otherwise, the test image will be classified as a nonface. In (5), $\hat{p}_{face/nonface}(\cdot \mid \cdot)$ stand for the estimated value of the local characteristics at site s based on the face and non-face training data bases, respectively. Recall that during the face detection phase, the pixels have already been permuted according to π_{opt}. Thus, the criteria in (5) implicitly depends on π_{opt}.

For face detection in test images with gray intensities ranging from 0-255, modulo 16 scaling converts the original intensities into the 0-15 range. Some blocking effect in the original image is observed after performing this step (see Figure 4). Automatic image scaling is carried out at several different scale values to detect faces of different sizes in the test image. However, multiscaling also increases the chances of false detection. A 20×15 window is moved in a raster scan fashion over the rescaled image. A gray level transformation is carried out for each window so that the mean and variance of gray intensities in the test window match that of the face training data base. This step is incorporated to detect relatively darker facial patterns. The values are calculated for each position of the detection window. If an LPR value is greater than 0, a face frame (white rectangular frame) is placed over the window. A post processing stage is also incorporated into the detection algorithm. Overlapping rectangular frames are merged together to form a rectangular frame that encompasses all the initial overlapping frames. Several threshold values, other than 0 (in (5)), are also considered. Possible faces correspond to high positive LPR values.

4.2. Face Synthesis

We use the estimated MRFs to synthesize both face and non-face examples. A Gibbs sampler was used to simulate examples from the respective MRFs in the following way: Start with an initial (random) realization of the grey level intensity values, say, $x^{(0)}$, for the entire image. At step n, $n = 1, 2, \ldots$, perform a sweep through all the sites in the image in a raster scan fashion. Denote by $x_s^{(n)}$ to be the value of the gray level intensity at site s during the n-th sweep of the raster scan. For each site s in S, generate a new value for x_s based on the conditional distribution specified in (3). Update $x_s^{(n)}$ to this new value, $x_s^{(n+1)}$. Move to a new site and repeat the above procedure again. The distribution of $x^{(n)} = (x_1^{(n)}, x_2^{(n)}, \ldots, x_{\#S}^{(n)})$ will approximately follow (2) for the π_{opt}-permuted sites after a large number of sweeps. Face and non-face examples are obtained by permuting the sites back to the natural order using the inverse permutation of π_{opt}.

Figure **??** and **??** each give examples of simulated faces and non-faces obtained from the above procedure for both the first order and second order MRF models. Each example in Figures **??** and **??** is an average of 10 images obtained by the simulation procedure described above, with a total of 500 sweeps for each image.

Subsequently, 10,000 simulated face and non-face images were obtained from the MRF models. These simulated examples were added to the existing training data base and the MRF models were re-trained as in Section 3; thus, we have increased the number of training samples in each data base from $N = 2000$ to $N = 12,000$. The re-trained models were used for face detection in test images using the LPR criteria stated in (5). Figures **??** give the results.

5. Summary and Conclusions

We have presented MRF models for face detection. The MRF models can be used for both synthesis and detection purposes. Synthesis can be used to increase the size of the training data base considerably. Re-training the MRF models for face detection in test images indicate significant reduction in false positive rates. Generally speaking, increasing the number of parameters of the MRF model (by increasing the order of the neighborhood system) results in more realistic simulated face images. Better detection properties are also obtained for the second order neighborhood system compared to the first order.

Acknowledgments

This research was supported in part by ONR grant no. N00014-01-1-0266.

References

[1] Y. Amit, D. Geman, and B. Jedynak. Efficient focusing and face detection. In H. Wechsler, editor, *Face Recognition: From Theory to Applications*, pages 143–158. NATO ASI Series F, Springer Verlag, Berlin, 1997.

[2] A. Colmenarez and T. Huang. Face detection with information-based maximum discrimination. *Proceedings of CVPR '97*, pages 782–787, 1997.

[3] S. C. Dass and A. K. Jain. Markov Face Models. *Proc. of the Eight International Conference on Computer Vision*, 2:680–687, 2001.

[4] N. Duta. *Learning based Detection, segmentation and matching of objects*. PhD thesis, Michigan State University, 2000.

[5] H. Rowley. *Neural Network-based Face Detection*. PhD thesis, Carnegie Mellon University, 1999.

[6] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(4):1019–1031, 1997.

[7] K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(1):39–52, 1998.