

Face Detection and Synthesis Using Markov Random Field Models *

Sarat C. Dass

Department of Statistics & Probability
Michigan State University
E Lansing, MI 48824, USA
sdass@stt.msu.edu

Anil K. Jain & Xiaoguang Lu

Department of Computer Science & Engineering
Michigan State University
E Lansing, MI 48824, USA
{jain, lvxiaogu}@cse.msu.edu

Abstract

Markov Random Fields (MRFs) are proposed as viable stochastic models for the spatial distribution of gray levels for images of human faces. These models are trained using data bases of face and non-face images. The trained MRF models are then used for detecting human faces in test images. We investigate the performance of the face detection algorithm for two classes of MRFs given by the first- and second-order neighborhood systems. From the cross validation results and from actual detection in real images, it is shown that the second-order model makes fewer false detections. We also investigate the possibility of increasing our training data base of faces by simulating face-like images from the trained MRFs. The performance of the re-trained MRFs based on added face-like images is compared to the original training data base.

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 approaches include face detection algorithm based on neural networks [8, 9], tree classifiers [1], distance from prototype criteria [10] and Markov Chains [6, 3]. However, the inherent spatial nature of digital images makes Markov Random Fields (MRFs) a natural choice for modeling the distribution of gray levels on face images. Preliminary results given in [5] indicate that the approach to detecting faces using MRFs appears promising. However, to circumvent the computational requirements, the MRFs used in [5] were only a valid approximation. In this paper, we develop exact estimation procedures that lead to parameter

estimates for the actual MRF model. Two classes of auto-MRF models, based on the first- and second-order neighborhood structures, are of interest. An immediate advantage of this stochastic approach over [5] 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. We investigate if subsequent re-estimation of model parameters and detection show a significant reduction of error rates for test images.

The MRF models used here do not utilize high level facial features 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 (e.g., eyes and mouth) 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 [5] where the most discriminating permutation of the sites is used compared to the natural order of site pixels.

2. Markov Random Field Models

Let $S = \{1, 2, \dots, M\}$ denote the collection of all sites in a $R \times C$ image, where $M = RC$. For each site s in S , we denote by x_s , the gray level at that site, $x_s \in [0, L - 1]$. Let x_{-s} be the gray levels of all sites in S excluding site s . The spatial distribution of gray levels, $\mathcal{X} = \{x_s, s \in S\}$ on S will be modeled as a MRF with an associated neighborhood system $\mathcal{N} = \{N_s, s \in S\}$, where N_s denotes the neighbors of site s . In Figure 1, $x_{s,2}, x_{s,4}, x_{s,7}$ and $x_{s,5}$ represent the north, west, south and east neighbors of x_s , respectively, for the first-order neighborhood system. For the second-order neighborhood system, the additional sites $x_{s,1}, x_{s,3}, x_{s,6}$ and $x_{s,8}$ are also taken to be neighbors of s .

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

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

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

Figure 1. First- and second-order neighborhood systems for the site s .

by

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

where

$$H(x_s | x_{-s}) = \alpha_s x_s + \sum_{t \in N_s} \beta_{st} x_s x_t, \quad (2)$$

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

$$p(\underline{x}) = \frac{\exp \left\{ \sum_s \alpha_s x_s + \sum_{s \sim t} \beta_{st} x_s x_t \right\}}{\sum_{x_1} \sum_{x_2} \dots \sum_{x_M} \exp \left\{ \sum_s \alpha_s x_s + \sum_{s \sim t} \beta_{st} x_s x_t \right\}}, \quad (3)$$

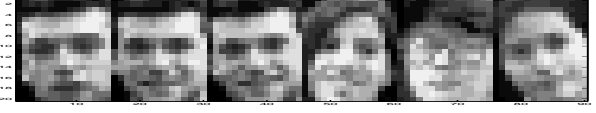
where $s \sim t$ stands for all pairs of sites s and t that are neighbors in S . The normalizing constant in (3) is hard to compute; thus, we resort to an approximation of the likelihood to avoid evaluating the normalizing constant. One such approximation is the pseudolikelihood (PL) defined by

$$PL = \prod_{s=1}^M \frac{\exp \left\{ \alpha_s x_s + \sum_{t \in N_s} \beta_{st} x_s x_t \right\}}{\sum_{x_s=0}^{L-1} \exp \left\{ \alpha_s x_s + \sum_{t \in N_s} \beta_{st} x_s x_t \right\}}. \quad (4)$$

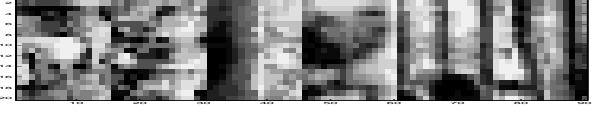
3. Model Training

The MRF models in Section 2 are fitted to a permutation π_{opt} of sites in the image which is obtained as follows. For a permutation π on $S = \{1, 2, \dots, M\}$, the gray intensity value at a site $s \in S$ is given by $x_{\pi(s)}$ (which is the gray value at the site $\pi(s)$). We choose the permutation π_{opt} that maximizes the Chi-square criteria given by

$$\chi^2(\pi) = \sum_{s=1}^N \frac{\{E_{face}(U_s^\pi) - E_{nonface}(U_s^\pi)\}^2}{E_{face}(U_s^\pi)} + \sum_{s=1}^N \frac{\{E_{face}(V_s^\pi) - E_{nonface}(V_s^\pi)\}^2}{E_{face}(V_s^\pi)}, \quad (5)$$



(a)



(b)

Figure 2. Examples of (a) faces and (b) non-faces in the training data base.

where $U_s^\pi = x_{\pi(s)}$, $V_s^\pi = \sum_{t \in N_s} x_{\pi(s)} x_{\pi(t)}$, and the expectations are computed based on the face and non-face images in the training data bases. The criteria in (5) avoids parameter estimation while finding the optimal permutation, π_{opt} , and incorporates information on spatial dependence. For the rest of this paper, the notation x_s will actually stand for $x_{\pi_{opt}(s)}$; we use the former notation for simplicity of exposition.

The MRF models 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 face). Each pixel can take 16 ($L = 16$) possible gray levels. The nonface examples are generated from images that resemble a face but are not actually so. The models were trained using 4,000 images from each of the face and non-face training data base. Figure 2 gives 6 examples of face and nonface images.

The value of the observed log pseudolikelihood (LPL) (see (4)) for the training sample is given as

$$LPL(\alpha, \beta) = \sum_{k=1}^N \sum_{s=1}^M \left(\alpha_s x_s^{(k)} + \sum_{t \in N_s} \beta_{st} x_s^{(k)} x_t^{(k)} - \log \left(\sum_{x_s=0}^{L-1} \exp \left\{ \alpha_s x_s + \sum_{t \in N_s} \beta_{st} x_s x_t^{(k)} \right\} \right) \right), \quad (6)$$

where $\alpha = \{\alpha_s, s \in S\}$, $\beta = \{\beta_{st}, s \in S, t \in N_s\}$, and $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_M^{(k)})$ are the observed gray levels of the k -th training image, $k = 1, 2, \dots, N$. We seek the maximum pseudolikelihood (MPL) estimates, $\hat{\alpha}$ and $\hat{\beta}$, that maximize the observed LPL . We assume toroidal (periodic) boundary conditions. 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 1,500. A multidimensional version

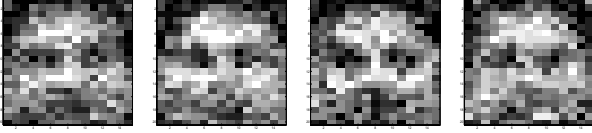


Figure 3. Simulated face-like images from the first-order model.

of the Newton-Raphson iteration procedure is used to find the MPL estimates of α and β .

4. Face Detection and Synthesis

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

$$LPR = \sum_{s=1}^M \log \left(\frac{\hat{p}_{face}(x_s^{test} | x_{-s}^{test})}{\hat{p}_{nonface}(x_s^{test} | x_{\pi-s}^{test})} \right) > 0. \quad (7)$$

Otherwise, the test image will be classified as a nonface. In (7), $\hat{p}_{face/nonface}(\cdot | \cdot)$ stands 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} . Hence, the criteria in (7) implicitly depends on π_{opt} .

For face detection in test images, modulo 16 scaling converts the original intensities into the 0-15 range. Automatic image scaling is carried out at several different scales to detect faces of different sizes. 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 levels match that of the face training data base. This step is incorporated to detect relatively darker facial patterns. The LPR 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 merges overlapping frames. Several threshold values, other than 0 (in (7)), are also considered.

Type I and Type II error rates for the first- and second-order MRFs models are estimated based on 4,000 face and 8,000 non-face images in the training data bases. The average Type I and Type II error rates (based on 5 runs) for the first-order (second-order) model are 0.060 (0.034) and 0.131 (0.064) with standard deviations 0.007 (0.006) and 0.010 (0.004), respectively. Clearly, the second-order MRF model performs better. The reported errors for the second order MRF model is also better than that of [5]. Figure 4 gives the results of the detection algorithm based on the second-order model for some independent test images. The results of the first-order model are not presented here.

Figure 3 gives several examples of simulated faces from a Gibbs sampler for the first-order MRF model; faces in Figure 3 are an average of 10 simulated images from the MRF model. The second-order model gives simulated images similar to those in Figure 3 but takes longer to generate. Subsequently, 4,000 simulated face images were added to the existing training data base and the first-order MRF model was re-trained as in Section 3, thus, increasing the number of training samples to 8,000. The cross validation procedure run on the re-trained first-order MRF model did not show any significant improvement over the original training data base.

5. Summary and Conclusions

We have presented MRF models for face detection and synthesis. Better detection results are obtained for the second-order model compared to the first-order. However, addition of simulated face images from the estimated MRF model do not significantly reduce detection error rates. There are several difficulties in fitting MRF models to face images, for example, due to pose. In future work, we would also like to incorporate facial features in model definition and fitting.

References

- [1] Y. Amit, D. Geman, and B. Jedynak. Efficient focusing and face detection. In H. Wechsler and J. Phillips, editors, *Face Recognition: From Theory to Applications*, pages 143–158. NATO ASI Series F, Springer Verlag, Berlin, 1997.
- [2] R. Chellappa and A. Jain, editors. *Markov Random Fields: Theory and Application*. Academic Press, Inc., 1991.
- [3] A. Colmenarez and T. Huang. Face detection with information-based maximum discrimination. *Proceedings of IEEE CVPR '97*, pages 782–787, 1997.
- [4] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 5:25–39, 1983.
- [5] S. C. Dass and A. K. Jain. Markov Face Models. *Proc. of IEEE ICCV '01*, 2:680–687, 2001.
- [6] N. Duta. *Learning based Detection, segmentation and matching of objects*. PhD thesis, Michigan State University, 2000.
- [7] S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Tokyo: Springer, 1995.
- [8] H. Rowley. *Neural Network-based Face Detection*. PhD thesis, Carnegie Mellon University, 1999.
- [9] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(4):1019–1031, 1997.
- [10] 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.



Image size: 300×310 . Detection time: 80 s.

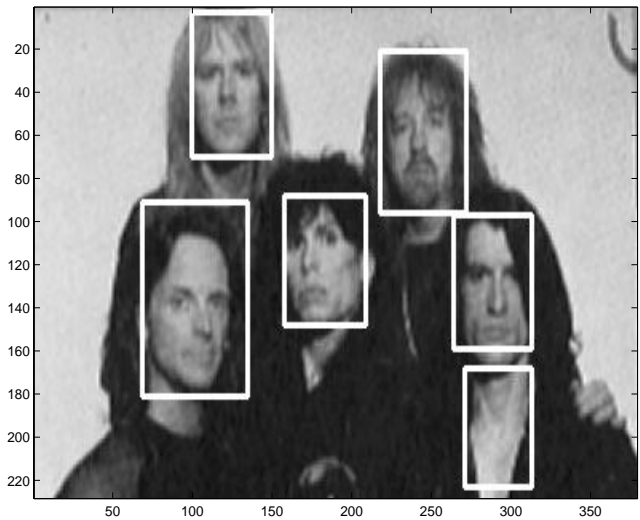


Image size: 230×400 . Detection time: 80 s.

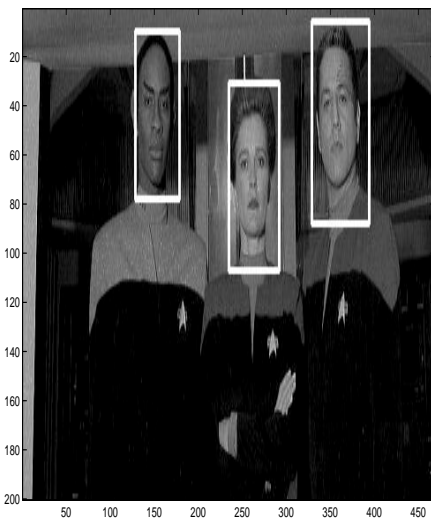


Image size: 200×460 . Detection time: 83 s.



Image size: 170×220 . Detection time: 33 s.

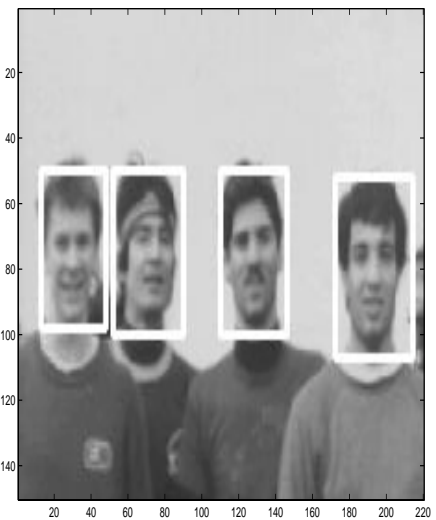


Image size: 150×230 . Detection time: 28 s.



Image size: 350×550 . Detection time: 185 s.

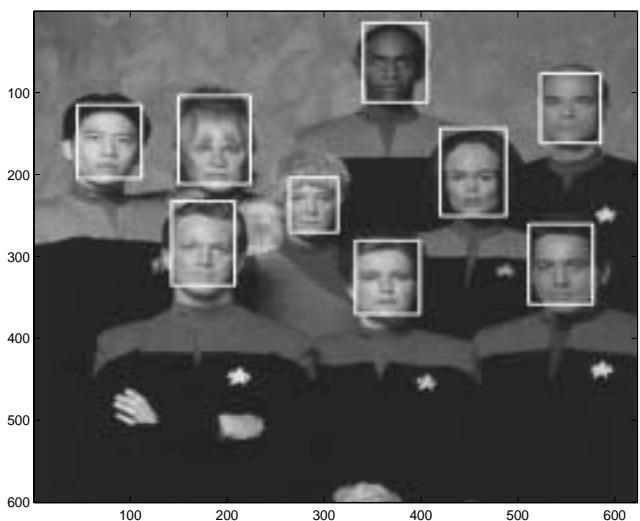


Image size: 600×620 . Detection time: 323 s.

Figure 4. Detection results based on the second-order model. The detection algorithm was written in C++ and run on a Pentium IV PC with 1.8GHz processing speed.