Markov Face Models

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

Anil K. Jain
Department of Computer Science & Engineering
Michigan State University
E Lansing, MI 48824, USA
jain@cse.msu.edu

Abstract

The spatial distribution of gray level intensities in an image can be naturally modeled using Markov Random Field (MRF) models. We develop and investigate the performance of face detection algorithms derived from MRF considerations. For enhanced detection, the MRF models are defined for every permutation of site indices (pixels) in the image. We find the optimal permutation that provides maximum discriminatory power to identify faces from nonfaces. The methodology presented here is a generalization of the face detection algorithm in [7, 5] where a most discriminating Markov chain model was used. The MRF models successfully detect faces in a number of test images.

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

1. Introduction

The use of Markov Random Fields (MRFs) to model spatial processes on lattices has been popular and widespread [3]. By using MRF models, one is able to model the behaviour of spatial processes locally via conditional distributions of attributes (gray values). In this paper, we use MRF automodels to represent the distribution of gray level intensities of facial images. Faces typically correspond to changes in gray level intensities along some spatial direction or at some special sites in the image. Our interest here is to determine whether MRFs capture these local changes in intensities for typical face images. There have been numerous attempts to detect faces in images using different techniques such as neural networks [15, 16], tree classifiers [2], distance from prototype criteria [17] and Markov Chains [7, 5]. Although Markov Chains use some notion of pixel dependence, this dependence is only allowed in one direction in space. For this reason, we feel that MRFs will be viable models for face detection since dependence can be captured along several spatial directions for different sites in the image. MRFs have also been successfully used for texture modeling, see [6], for example. Since facial images can be viewed as a type of texture in some sense, this is another reason we feel that MRFs will also be good models for facial 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.

In order to achieve better detection rates, we seek an optimal permutation of sites in the image for which the MRF model has the best fit. In other words, for detection between faces and nonfaces, it can turn out that a permutation of the sites in the image has better discriminatory power to distinguish between a face and a nonface compared to the original (unpermuted) sites. We call the resulting MRF the most discriminating MRF for detecting faces. Thus, the most discriminating MRF approach is a generalization of the most discriminating Markov Chain approach of [7, 5].

It is essential that low level detection algorithms be computationally efficient. For the most discriminating Markov Chain approach, this is definitely the case since sites can be updated sequentially utilizing the Markovian structure. It is well known that the normalizing constant in MRF models causes great difficulty in computations and may actually compromise the efficiency of the algorithm. For this reason, we avoid likelihoods resulting from MRF models. Instead, we use pseudolikelihoods and pseudolikelihood ratios for estimating model parameters and for subsequent detection. The resulting reduction in computational time and complexity is significant.

The remainder of this paper is organized as follows. In Section 2, we present the basic MRF models that we use. In Section 3, we discuss the procedures to train and cross

	x_{sn}	
x_{sw}	x_s	x_{se}
	x_{ss}	

Figure 1. First order neighbors of site s and corresponding gray level intensities.

validate the MRF models. We present MRFs defined via a permutation of sites in Section 4. The Chi-square criteria to find an optimal permutation of sites for face detection is given. We also present cross validation results of our detection algorithm in Section 4. Finally, the performance of our detection algorithm on real images is illustrated in Section 5.

2. Markov Random Field Models

Let $S = \{1, 2, \cdots, N\}$ denote the collection of all sites in a $R \times C$ image, where $N = RC$. For each site s in S, we denote by x_s 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} to be the gray level intensities of all sites in S excluding site s. 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). For any MRF, there is an associated neighborhood system $\mathcal{N} = \{N_s, s \in S\}$, where N_s denotes the neighbors of site s. We consider only the first order neighborhood structure for the MRF models (see Figure 1). In Figure 1, x_{sn}, x_{sw}, x_{ss} and x_{se} represent the north, west, south and east neighbors of x_s, respectively. For a MRF, the local characteristic (conditional distribution) of x_s given x_{-s} at each site s depends only on the neighbors of s, namely, $\{x_t, t \in N_s\}$.

Markovian models are, in general, parameterized by a certain number of coefficients which govern the degree of spatial correlation between sites. These coefficients are unknown in typical applications and have to be estimated from training samples. It is well known that there is a trade off between the number of parameters to be estimated and reliability of the overall model fit to the data. We present a general class of MRFs for faces in this paper and focus our attention on two special cases. The two special cases of MRFs are characterized by different number of parameters (3 and 468, respectively). We investigate the overall fit of these two classes of models for faces and non-faces in the training samples. Finally, the trained models are used to detect faces in independent test images.

2.1. Automodels

We consider the general class of auto MRF models (see [8]) as candidate models for faces. The local characteristics at site s for auto MRF models is given by

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

where $H(x_s \mid x_{-s}) = \alpha_s x_s + \sum_{t \in N_s} \beta_{st} x_s x_t$ represents the conditional potential function corresponding to the local characteristics at site s. The conditional specifications in (1) are parameterized by the coefficients $\{\alpha_s, s \in S\}$ and $\{\beta_{st}, s \in S, t \in N_s\}$. It follows from Brooks' expansion that these conditional distributions uniquely determine a joint distribution (likelihood) on S (provided $\beta_{st} = \beta_{ts}$) given by

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

where $s \sim t$ stands for all pairs of sites s and t that are neighbors in S. The likelihood in (2) is hard to maximize with respect to the parameters $\{\alpha_s\}$ and $\{\beta_{st}, s \in S, t \in N_s\}$ due to the presence of the normalizing constant in the denominator. Approximations to the likelihood, such as the pseudolikelihood (PL), have been considered to avoid handling of the normalizing constant. The pseudolikelihood (PL) for the MRF automodels specified by the conditional distributions in (1) is given by

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

We consider two special cases of the auto MRF models in (2). The first class of models, henceforth called Model I, is obtained by taking $\alpha_s = \alpha$, $\beta_{st} = \beta_h$ if $t \in \{w, e\}$ and $\beta_{st} = \beta_v$ if $t \in \{n, s\}$ (see Figure 1) in (2). In other words, Model I captures overall gray level intensity (via α) and spatial correlations, via β_h and β_v, in the horizontal and vertical directions, respectively. In this case, the joint

distribution can be written as

$$p(\underline{x}) = \frac{\exp\left\{\alpha T_{overall} + \sum_d \beta_d T_d\right\}}{\sum_{x_1=0}^{L-1}\sum_{x_2=0}^{L-1}\cdots\sum_{x_N=0}^{L-1}\exp\left\{\alpha T_{overall} + \sum_d \beta_d T_d\right\}} \quad (4)$$

where the sum over d represents the sum over vertical and horizontal directions ($d = \{v, h\}$, respectively), and

$$T_{overall} = \sum_s x_s \quad \text{and} \quad T_d = \sum_s \sum_{s \overset{d}{\sim} t} x_s x_t,$$

where the sum $s \overset{d}{\sim} t$ is taken over all neighbors of s in the direction $d, d = \{v, h\}$. The pseudolikelihood for Model I is given by

$$PL(\text{Model I}) = \prod_{s=1}^{N} \frac{\exp\left\{\alpha x_s + \sum_d \beta_d \sum_{s \overset{d}{\sim} t} x_s x_t\right\}}{\sum_{x_s=0}^{L-1}\exp\left\{\alpha x + \sum_d \beta_d \sum_{s \overset{d}{\sim} t} x_s x_t\right\}}. \quad (5)$$

Since α, β_h and β_v are independent of s, Model I assumes that all sites have the same degree of spatial correlation.

Our main reason for considering the auto MRF model in (1) is to determine if the information present in special sites (for example, the locations of eyes, nose and facial outline) is actually used by the MRF when distinguishing between a face and nonface. The importance of a site can be determined by relative magnitudes of the site coefficients, (α_s and β_{st}, $t \in N_s, s \in S$) in typical face and nonface images. However, estimating the coefficients in (2) is computationally challenging because of difficulty in handling the normalizing constant. This problem is not alleviated when the pseudolikelihood in (3) is used since each parameter β_{st}, with $\beta_{st} = \beta_{ts}$, occurs in the conditional specifications of more than one site.

Thus, we consider another class of models, henceforth called Model II, by considering a simple approximation to (3) while retaining the ability to measure the importance of special sites. For each site s, we consider a parameter β_s that measures the overall importance of $\{\beta_{st}, t \in N_s\}$ in a face image. Instead of (3), we consider the following approximation for the pseudolikelihood

$$PL(\text{Model II}) = \prod_{s=1}^{N} \frac{\exp\{\alpha_s U_s + \beta_s V_s\}}{\sum_{x_s=0}^{L-1}\exp\left\{\alpha_s x_s + \beta_s \sum_{t \in N_s} x_s x_t\right\}}, \quad (6)$$

where $U_s = x_s$ and $V_s = \sum_{t \in N_s} x_s x_t$ represent the gray level intensity of pixel s and the joint moment of neighboring gray level intensities, respectively. The pseudolikelihood in (6) is obtained by taking $\beta_{st} = \beta_s$ for $t \in N_s$ in (3). For each site s, β_s measures the "average" correlation of x_s with its neighbors. However, since the condition $\beta_{st} = \beta_{ts}$ is not satisfied for neighboring sites s and t, Model II does not define a proper joint distribution for the gray levels at all sites in the image (as in (2)). Nevertheless, the approximate MRF model can assess the relative importance of site s via α_s and β_s in discriminating between a face and a nonface. Thus, we use Model II in the context of face detection only and not for the purpose of simulating facial patterns. The parameters $\{\alpha_s, \beta_s\}$ in (6) can be estimated separately for each site s which entails great reduction in computational complexity. This is not possible for (3).

3. Training the MRF Models and Cross Validation Results

The MRF models (Models I and II) given 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 training database consists of 7,200 and 8,422 images (of size 20×15) of faces and nonfaces, respectively. Figures 2 and 3 each give 6 examples of face and nonface images in the training database. We fit each class of models for faces and nonfaces training samples. We estimate the unknown parameters in each model by the Maximum Pseudolikelihood (MPL) method, that is, by maximizing the pseudolikelihoods given in (5) and (6), with respect to the unknown parameters.

3.1. Detection Algorithm

Subsequent to parameter estimation, the trained models are used for face detection. We classify an input image, $\{x_s^{inp}, s \in S\}$, as a face if the log pseudolikelihood ratio,

$$LPR = \log\left(\prod_{s=1}^{N} \frac{\hat{p}_{face}(x_s^{inp} \mid x_{-s}^{inp})}{\hat{p}_{nonface}(x_s^{inp} \mid x_{-s}^{inp})}\right) \quad (7)$$

$$= \sum_{s=1}^{N} \log\left(\frac{\hat{p}_{face}(x_s^{inp} \mid x_{-s}^{inp})}{\hat{p}_{nonface}(x_s^{inp} \mid x_{-s}^{inp})}\right) \quad (8)$$

$$> 0. \quad (9)$$

Otherwise, the input image will be classified as a nonface. In (7) and (8), $\hat{p}_M(\cdot \mid \cdot)$ stands for the estimated value of

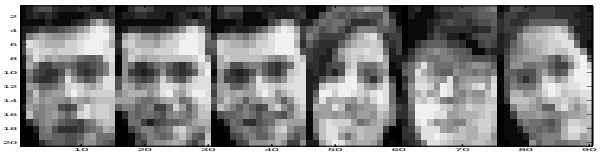


Figure 2. Examples of faces in the training data (20×15 images with 16 gray levels).

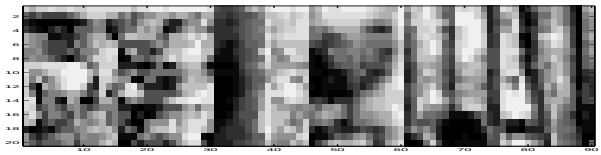


Figure 3. Examples of nonfaces in the training data (20×15 images with 16 gray levels)**.**

the local characteristics at site s after the parameters have been estimated from each $M = \{face, nonface\}$ training data sets. The criteria stated in (9) is in terms of the sum of logarithms of pseudolikelihood ratios for faces and nonfaces, and will be called the log pseudolikehood ratio (LPR) criteria.

3.2. Cross Validation Results

We use cross validation to obtain estimates of Type I and Type II errors for Models I and II. Type I error is viewed as the more costly of the two, since a post processing stage which detects facial features can eliminate most of the falsely detected faces [9]. Both the training data sets of faces and nonfaces are randomly divided into two groups; the first group for model training and parameter estimation, and the second group for estimation of misclassification probabilities using the LPR criteria. For Model I, 5 runs of the cross validation procedure gave an average Type I error of 0.3667 with an estimated standard deviation of 0.03. For Model II, the results of the cross validation procedure are given in the second and third columns of Table 1. The superior performance of Model II compared to Model I is due to the ability of Model II to capture site importance; the degree of spatial association in face images varies significantly for different sites in the image. Since we found Model II to be more satisfactory for face detection, subsequent investigations were carried out for Model II only.

Denote by f and g to be the density estimates corresponding to the distribution of LPR values (see (7) and (8)) under the face and nonface images, respectively, obtained from the cross validation procedure. A measure of the degree of overlap between f and g is given by

$$D(f,g) = \int_R \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx.$$

It can be shown that $0 \leq D(f,g) \leq 2$, with $D(f,g) = 0$ iff $f = g$, and $D(f,g) = 2$ if f and g are completely disjoint. Small values of D in the fourth column of Table 1 indicate the the distributions of face and nonface are not very well separated.

4. Most Discriminating MRF Models via Permutations

For better detection purposes, we investigate if the MRF models are a better fit to a *permutation* of the sites in the image, instead of the natural ordering of the sites. We consider the class of all permutations of sites 1 to N, and choose that permutation which gives maximum discriminatory power for detecting faces. One argument for considering permutations of site indices is that the joint association of x_{π_s} and x_{π_t}, for a permutation π, may be better at discriminating between faces and nonfaces compared to x_s and x_t. Thus, following the construction of joint MRF models on S using the conditional specifications in Section 2, one can similarly define local characteristics for a given permutation π, namely,

$$p(x_{\pi_s} \mid x_{\pi_{-s}}) = \frac{\exp\{H(x_{\pi_s} \mid x_{\pi_{-s}})\}}{\sum_{x_{\pi_s}=0}^{L-1} \exp\{H(x_{\pi_s} \mid x_{\pi_{-s}})\}}, \quad (10)$$

where $H(x_{\pi_s} \mid x_{\pi_{-s}}) = \alpha_s x_{\pi_s} + \sum_{t \in N_s} \beta_{st} x_{\pi_s} x_{\pi_t}$.

The joint MRF model specified by the local characteristics in (10) becomes

$$p(\underline{x}) = \frac{\exp\left\{\sum_s \alpha_s x_{\pi_s} + \sum_{s \sim t} \beta_{st} x_{\pi_s} x_{\pi_t}\right\}}{\sum_{x_1}\sum_{x_2}\cdots\sum_{x_N} \exp\left\{\sum_s \alpha_s x_{\pi_s} + \sum_{s \sim t} \beta_{st} x_{\pi_s} x_{\pi_t}\right\}}. \quad (11)$$

For Model II, the approximate pseudolikelihood (PL) is given by

$$PL(\pi) = \prod_{s=1}^{N} \frac{\exp\{\alpha_s U_s^\pi + \beta_s V_s^\pi\}}{\sum_{x_{\pi_s}=0}^{L-1} \exp\left\{\alpha_s x_{\pi_s} + \beta_s \sum_{t \in N_s} x_{\pi_s} x_{\pi_t}\right\}}. \quad (12)$$

where $U_s^\pi = x_{\pi_s}$ and $V_s^\pi = \sum_{t \in N_s} x_{\pi_s} x_{\pi_t}$ are the counterparts of U_s and V_s in Section 2 for a given permutation π.

Table 1. Crossvalidation results for Model II (natural order of sites)

Run No.	Type I Error	Type II Error	D
1	0.1587	0.1007	0.92
2	0.1553	0.1067	0.91
3	0.1753	0.0960	0.93
4	0.1573	0.0987	0.94
5	0.1420	0.1080	0.91

Table 2. Crossvalidation results for Model II (optimal permutation of sites)

Run No.	Type I Error	Type II Error	D
1	0.0920	0.0787	1.24
2	0.0947	0.0767	1.22
3	0.1027	0.0780	1.20
4	0.1060	0.0773	1.22
5	0.0907	0.0767	1.26

4.1. Chi-square criteria for Model II

We seek a criteria that will maximize discrimination between face and nonface images based on a permutation of sites. In [7] and [5], the Kullback-Leibler distance (which involves the likelihoods of face and nonface images) was used in the case of Markov chain models. Using the likelihoods in the discrimination criteria is not feasible in the case of MRF models, since the normalizing constants cannot be broken down in simpler sum components as was done in the case of the Markov chain model. Using the ratio of pseudolikelihoods is easier compared to the full likelihood but computing the discrimination criteria based on pseudolikelihoods is still time consuming. Therefore, we resort to a Chi-square discrimination criteria that is discussed below for MRFs.

We propose a criteria based on the chi-square statistic to obtain a permutation, π^{opt}, that best distinguishes between a face and a non-face. For the approximate PL in (12), the relevant site statistics are given by U_s^π and V_s^π for each site s. We use the following Chi-square criteria for discrimination

$$\chi^2(\pi) = \sum_{s=1}^{N} \frac{\{E_{face}(U_s^\pi) - E_{nonface}(U_s^\pi)\}^2}{E_{face}(U_s^\pi)} + \sum_{s=1}^{N} \frac{\{E_{face}(V_s^\pi) - E_{nonface}(V_s^\pi)\}^2}{E_{face}(V_s^\pi)}, \quad (13)$$

where the expectations are computed under the (unknown) spatial distributions of the face and non-face images. Thus, the expected values in (13) are estimated using the sample averages from the face and nonface training data set. Denote the total number of face and nonface images in the training data set by N_f ($N_f = 7,200$) and N_{nf} ($N_{nf} = 8,422$), respectively. For every permutation π, the estimates of $E_M(U_s^\pi)$ and $E_M(V_s^\pi)$ are

$$\hat{E}_M(U_s^\pi) = \frac{1}{N_M} \sum_{k=1}^{N_M} x_{\pi_s}^{(k)}$$

and

$$\hat{E}_M(V_s^\pi) = \frac{1}{N_M} \sum_{k=1}^{N_M} \sum_{t \in N_s} x_{\pi_s}^{(k)} x_{\pi_t}^{(k)}$$

for $M = \{face, nonface\}$ and $N_M = \{N_f, N_{nf}\}$, respectively.

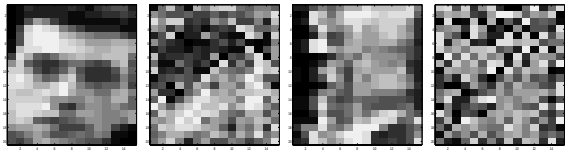
4.2. Finding the Best Permutation using the Chi-square Criteria

Since the space of all permutations is extremely large, ($\mathcal{O}(N!)$, for N sites), we resort to simulated annealing (SA) to find the best permutation according to (13) for Model II. The SA algorithm [1] is described as follows. Start with an initial permutation, π_0, and initial temperature, $T = t_0$, say. Randomly select two sites for interchange and obtain the updated permutation, π_1. For π_1, calculate the Chi-square distance between faces and nonfaces in the training set. If this distance is larger than the initial Chi-square distance for π_0, accept the new permutation, π_1. Otherwise, accept the new permutation, π_1, with probability e^δ, where δ is the difference between the two chi-square values (computed according to (13)) for π_1 and π_0. The acceptance-rejection scheme is carried out for a large number of runs. Subsequently, T is reduced to, say, t_1, and the above algorithm is repeated for the temperature, t_1. The SA procedure reaches a solution that is close to the global optimal solution when T is small. The acceptance-rejection scheme for each temperature level was carried out for $n = 1000$ times. The cooling schedule was taken to be $T = T * 0.97$, with $t_0 = 5$.

Once the best permutation was found, the parameters of the MRF for faces and nonfaces were estimated using the Maximum Pseudolikelihood (MPL) method.

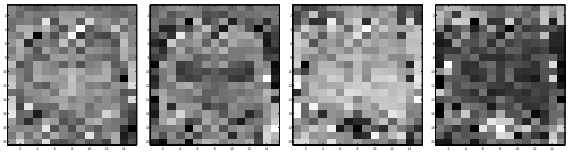
4.3. Detection Algorithm

For the optimal permutation, π^{opt}, and the corresponding estimated parameters (for both the face and nonface MRF models), a input image, $\{x_s^{inp}, s \in S\}$, is classified as a



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

Figure 4. Permutations of sites. (a) Face example, (b) Permuted face, (c) Nonface example, (d) Permuted nonface.



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

Figure 5. Parameter values for faces and nonfaces. (a) $\{\alpha_s\}$ for faces, (b) $\{\beta_s\}$ for faces, (c) $\{\alpha_s\}$ for nonfaces, (d) $\{\beta_s\}$ for nonfaces.

face if

$$\sum_{s=1}^{N} \log \left(\frac{\hat{p}_{face}(x^{inp}_{\pi^{opt}_s} \mid x^{inp}_{\pi^{opt}_{-s}})}{\hat{p}_{nonface}(x^{inp}_{\pi^{opt}_s} \mid x^{inp}_{\pi^{opt}_{-s}})} \right) > 0. \qquad (14)$$

Otherwise, the input image will be classified as a nonface. In (14), $\hat{p}_M(\cdot \mid \cdot)$ stands for the estimated value of the local characteristics at site s after the optimal permutation π^{opt} has been found, and the parameters have been estimated under $M = \{face, nonface\}$, respectively. This is again the log pseudolikelihood ratio (LPR) criteria for the permuted sites.

4.4. Cross Validation Results

The results of the cross validation procedure for permuted sites are given in Table 2. The cross validation procedure is run 5 times. It is clear from Tables 1 and 2 that the permuted Model II has better detection property compared to the unpermuted sites. This can also be seen from the increase in the values of D in Table 2.

It is interesting to see how the optimal permutation for Model II rearranges gray level intensities in an image. Figure 4 (a) shows a typical face image from the training data base. The optimal permutation is applied to the face image and the resulting image is presented in Figure 4 (b). It is clear that the optimal permutation forms two distinct clusters of gray level intensities, one cluster of low gray level intensities while another cluster of higher gray level intensities. The relative positions of these clusters in a face image are also fixed for different face images. No such cluster are formed when a nonface image is considered. See Figures 4 (c) and 4 (d), for example.

We also display the site coefficients, $\{\alpha_s\}$ and $\{\beta_s\}$, of Model II for faces and nonfaces. The image plots are obtained first by rescaling the coefficients to the $0-255$ range, and then reordering the permuted sites back to the natural order. Figures 5 (a) and (b) show the relative magnitude of the $\{\alpha_s\}$ and $\{\beta_s\}$, respectively, for a face image. Observe that the $\{\beta_s\}$ image extracts the distinguishing features of a face, namely the face outline, and the positions of the eyes and nose. Since the eyes and nose are relatively darker regions compared to the surrounding sites, $\{\beta_s\}$ at the boundaries of the eyes, nose and face outlines capture this change in gray level intensities. Since the intensities change in opposite directions (from lighter to darker, or vice versa), this is reflected in the $\{\beta_s\}$ coefficients by their large negative values.

Original image | Modulo 16 reduction | Scaled Image

Figure 6. Effects of Blocking and Scaling.

5. Experimental Results

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 6). In order to fit a face in the test images into our 20×15 detection frame, we scale (up or down) the original image so that the faces approximately fit into the detection frame; we assume that the scaling factor is known for each test image considered. 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 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 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 (14)), are also considered. Possible faces correspond to high positive LPR values.

We applied our algorithm on 102 test images each containing a single frontal face view from the FERET database [14]. All 102 faces were detected with 13 false alarms. Figure 7 shows the performance of the face detection algorithm applied to a number of test images containing multiple faces.

The detection algorithm was written in MATLAB and was run on a PC with a 750 Mhz Pentium III processor. The detection times (in seconds) for these images are approximately 170s for (a)-(c), 80s for (d), 300s for (e) and 330s for (f). Detection times can be improved considerably by coding the algorithm in a compilable language such as C.

6. Summary and Conclusions

We have presented MRF models for face detection. Better detection properties are obtained for a permutation of the sites, instead of the natural ordering. Models that account for varying degrees of spatial correlation according to sites are better as face models. Future work includes investigating multiple MRF models for faces and extensions to color images. Automatic scaling factors that work for a variety of test images will also be incorporated into our detection algorithm.

Acknowledgments

The authors would like to thank Nico Duta and Vincent Hsu for helpful discussions and for making the face/nonface databases available to the authors. This research was supported in part by ONR grant no. N00014-01-1-0266.

References

[1] E. Aarts and J. Korst. *Simulated Annealing and Boltzmann Machines: a Stochastic Approach to Combinatorial Optimization and Neural Computing*. Wiley, Chichester, 1989.

[2] 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.

[3] R. Chellappa and A. Jain, editors. *Markov Random Fields: Theory and Application.* Academic Press, Inc., 1991.

[4] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces: a survey. *Proc. IEEE*, 83:705–740, 1995.

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

[6] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE Trans. Pattern Analysis and Machine Intelligence.*, 5:25–39, 1983.

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

[8] X. Guyon. *Random Fields on a Network*. Springer-Verlag, 1995.

[9] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face detection in color images. *Proc. IEEE Int'l Conference Image Processing*, Oct 2001. To appear.

[10] K. M. Lam and H. Yan. Locating and extracting the eye in human face images. *Pattern Recognition*, 29(5):771–779, 1996.

[11] D. Maio and D. Maltoni. Real-time face location on gray-scale static images. *Pattern Recognition*, 33(9):1525–1539, Sept 2000.

[12] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pages 130–136, June 1997.

[13] P. S. Penev and J. J. Artick. Local feature analysis: A general statistical theory for object representation. *Network: Computation in Neural Systems*, 7(3):477–500, 1996.

[14] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-detection algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22:1090–1104, Oct 2000.

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

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

[17] 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.

[18] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[19] H. Wechsler, P. Phillips, V. Bruce, and T. Huang, editors. *Face Recognition: From Theory to Applications*. Springer-Verlag, 1998.

[20] L. Wiskot, J. M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.

[21] K. C. Yow and R. Cipolla. Information theory and face detection. *IEEE Int'l Conf. Pattern Recognition*, pages 601–605, 1996.

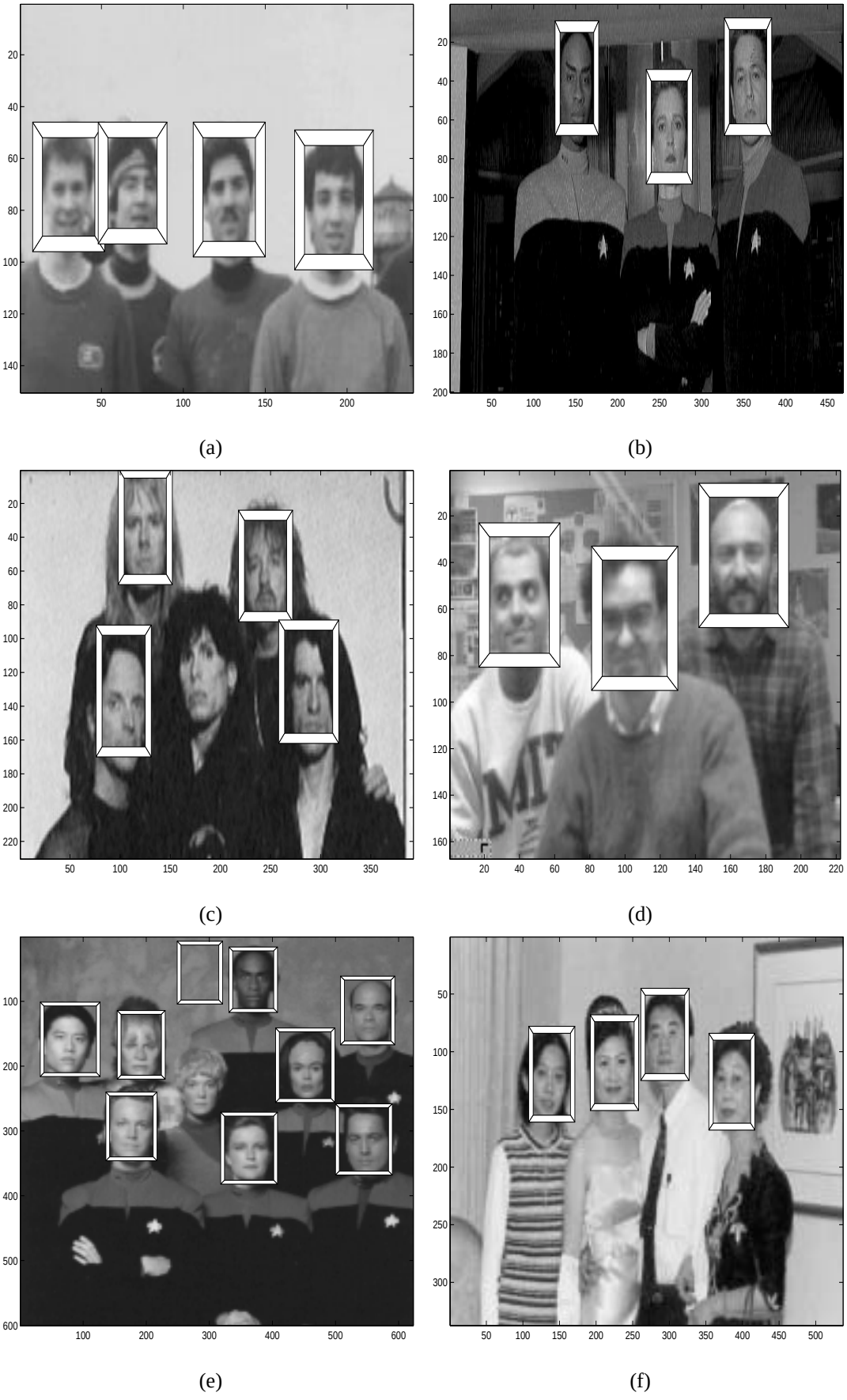


Figure 7. Input images, sizes (a) 150×250, (b) 200×460, (c) 230×400, (d) 170×220, (e) 600×620, (f) 350×550.