Fig. 2.   Bayes error estimates for SONAR data transformed to 10-dimensional space by Patrick–Fisher's algorithm (solid line) and E (dotted line).

high-dimensional data these results might be more in favor of E.) This is a result of the fact that each iteration of simplex requires that the samples be transformed to the low-dimensional space, and then the Bayes error estimated in that space, which in turn requires computation of distances, determination of optimal thresholds, and classification. The iterative process is repeated until the convergence is achieved. In contrast, FLD and KL are noniterative techniques, which merely compute eigenvalues and eigenvectors of certain matrices. For PF, which is also an iterative algorithm, the difference in CPU timing is about an order of magnitude for SONAR data.

V. Summary

The aim of this correspondence was to investigate the possibility of constructing such linear transformation of labeled multidimensional vectors that would hopefully ensure the maximum attainable classification accuracy in the transformed space. In other words, the goal was to come as close as possible to computing a transformation matrix that would minimize the Bayes error in the low-dimensional space, and to devise a practical algorithm for such purpose. The proposed algorithm, called E, consists in finding such matrix that minimizes the estimate of the Bayes error, computed on the training data set projected to the low-dimensional space. The most reliable technique for Bayes error estimation available was used, and the simplex algorithm played the role of the optimization algorithm.

In all examples, E demonstrated superior performance in comparison with standard algorithms, coming close to the theoretical limits on classification accuracy. This is payed through significant increase of the computational load. Still we managed to keep its complexity within realistic bounds, thus realizing designated goals.

Acknowledgment

References

[1] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 8, pp. 308–313, 1965.

[2] L. F. Guseman, Jr., B. C. Peters, Jr., and H. F. Walker, "On minimizing the probability of misclassification for linear feature selection," *Annals Statist.*, vol. 3, pp. 661–668, 1975.

[3] Š. A. Dudani, K. J. Breeding, and R. B. McGhee, "Aircraft identification by moment invariants," *IEEE Trans. Comput.*, vol. C-26, pp. 39–45, 1977.

[4] A. K. Jain and W. G. Waller, "On the optimal number of features in the classification of multivariate Gaussian data," *Pattern Recogn.*, vol. 10, pp. 365–374, 1978.

[5] S. J. Raudys, "Determination of optimal dimensionality in statistical pattern classification," *Pattern Recogn.*, vol. 11, pp. 263–270, 1979.

[6] G. Biswas, A. K. Jain, and R. C. Dubes, "Evaluation of projection algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, pp. 701–708, 1981.

[7] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach.* Englewood Cliffs, NJ: Prentice-Hall International, 1982.

[8] J. Kittler, "Feature selection and extraction," in *Handbook of Pattern Recognition and Image Processing*, T. Y. Young and K.-S. Fu, Eds. Orlando, FL: Academic, 1986.

[9] K. Fukunaga and D. M. Hummels, "Bayes error estimation using Parzen and k-NN procedures," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 634–643, 1987.

[10] R. P. Gorman and T. J. Sejnowski, "Analysis of hidden units in a layered network trained to classify sonar targets," *Neural Networks*, vol. 1, pp. 75–89, 1988.

[11] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. Boston: Academic, 1990.

[12] P. Pejnović, "Application of moment invariants in visual pattern recognition" (in Serbo-Croatian), M.S. thesis, Faculty of Electrical Engineering, Univ. of Belgrade, Belgrade, Yugoslavia, 1991.

[13] Lj. Buturović and Z. Stojiljković, "PARIS—pattern analysis and recognition interactive system," *Pattern Recogn.*, vol. 24, p. 11, 1991.

[14] Lj. J. Buturović, "On the minimal dimension of sufficient statistics," *IEEE Trans. Inform. Theory*, vol. 38, pp. 182–186, 1992.

On a Parameter Estimation Method For Gibbs–Markov Random Fields

Mehmet İ. Gürelli and Levent Onural

Abstract—This correspondence is about a Gibbs–Markov random field (GMRF) parameter estimation technique proposed by Derin and Elliott. We will refer to this technique as the histogramming (H) method. First, the relation of the H method to the (conditional) maximum likelihood (ML) method is considered. Second, a bias-reduction based modification of the H method is proposed to improve its performance, especially in the case of small amounts of image data.

Index Terms—Image modeling, texture, Gibbs–Markov random fields, parameter estimation, pattern recognition.

I. Introduction

Texture plays a very important role in image processing. In the literature, there exist several techniques for the analysis and processing of textured images, such as classification and identification [1], and segmentation and restoration [2]–[9]. Among the most well-known texture models are the autoregressive models [2], random mosaic models [10], and stochastic models [3]–[9], [11]–[14]. The Gibbs

M. İ. Gürelli was with the Department of Electrical and Electronics Engineering, Bilkent University, 06533 Ankara, Turkey. He is now with the Signal and Image Processing Institute, Department of Electrical Engineering—Systems, University of Southern California, Los Angeles, CA 90089.

L. Onural is with the Department of Electrical and Electronics Engineering, Bilkent University, 06533 Ankara, Turkey.

IEEE Log Number 9214458.

and Markov random fields are among the most powerful stochastic models [11]–[14]. In this correspondence, we will collectively refer to both models as the Gibbs–Markov random field (GMRF) model. Several segmentation and restoration algorithms based on the GMRF model can be found in [4]–[9].

In many applications of the GMRF texture model, it is usually necessary to estimate the model parameters (see, for example, [7]). The *coding* and *pseudo-likelihood* methods are among the most widely known parameter estimators [12], [13]. The pseudo-likelihood method is an extension of the coding method from a single coding to the whole image region, disregarding the lack of conditional independence among the pixel values. In this correspondence, the coding and pseudo-likelihood methods will be collectively referred to as the *maximum likelihood* (ML) method. Another method, which we will call as the *histogramming* (H) method, has been proposed by Derin and Elliott [7].

In this correspondence, we concentrated on the H method. Our aim is twofold. First, we want to clarify the properties of the H method and establish its relation to the well-known ML method. Second, we propose a modification to the H method to improve its performance for binary GMRF's, especially in the case of small amounts of image data. In the case of small amounts of image data, the H method suffers from two facts. First, the bias in parameter estimates becomes more emphasized, and second, in the H method, some data is generally discarded, which usually results in an insufficient amount of data to obtain the GMRF parameter estimates uniquely. As will be clarified in the forthcoming sections, the modification we proposed in this paper addresses these two problems simultaneously.

In Section II mathematical preliminaries are given. In Section III, the H method is derived. Section IV is aimed at clarifying the link of the H method to the ML method. In Section V, the bias in the H method is described. In Section VI, a method to reduce the bias is proposed. In Section VII, we give some experimental results, and, finally, some conclusions are given in Section VIII.

II. MATHEMATICAL PRELIMINARIES

We model the image region as a set of rectangularly placed pixels denoted by $\mathcal{D}$. We will refer to the numerical realization of the random field at pixel $i \in \mathcal{D}$ by the notation $y(i)$. For the numerical realization over a subset $\mathcal{R} \subset \mathcal{D}$, the notation $\mathcal{Y}(\mathcal{R})$ will be used. Detailed descriptions of GMRF's and related concepts such as *neighborhood systems* and *codings* can be found in [7] and [11]–[13]. The shape of the neighborhood of an interior point i of $\mathcal{D}$ will be assumed to be independent of its location in $\mathcal{D}$ and will be denoted by η_i (or simply η). For a given coding $\mathcal{C}$, GMRF's have the following (conditional) independence property [11]:

$$P(\mathcal{Y}(\mathcal{C})|\mathcal{Y}(\mathcal{D} \setminus \mathcal{C})) = \prod_{i \in \mathcal{C}} P(y(i)|\mathcal{Y}(\eta_i)). \tag{1}$$

This property is very important in the formulation of the H and ML methods. To keep the discussions clear, we will refer to a binary GMRF described by the conditional probability distribution [11].

$$P(y(i) = s|\mathcal{Y}(\eta_i)) = \frac{e^{sT}}{1 + e^T} \tag{2}$$

where $T = \alpha + \beta_h(u+u') + \beta_c(v+v') + \beta_m(t+t') + \beta_r(w+w')$ (3)

and $s, u, u', v, v', t, t', w, w' \in \{0,1\}$ are the pixel values whose relative locations are as shown in Fig. 1. The vector of model parameters $\underline{\omega} = [\alpha \ \beta_h \ \beta_c \ \beta_m \ \beta_r]^T$ will be called the GMRF parameter vector. Multilevel GMRF's of the above parametric form may be found in [7] and [11].

Fig. 1. Neighborhood pixel values.

III. DERIVATION OF THE HISTOGRAMMING (H) METHOD

The GMRF parameter estimation method described in this section has been proposed by Derin and Elliott [7]. We will refer to this method as the *histogramming* (H) method. We start by partitioning the set of possible neighborhood realizations into K groups such that within the kth group ($k = 1, \cdots, K$) each neighborhood realization yields the same coefficient vector $\underline{x}_k$, given as

$$\underline{x}_k = [1 \ (u + u') \ (v + v') \ (t + t') \ (w + w')]^T. \tag{4}$$

A possible grouping is to assume that each neighborhood realization is a distinct group. We will indicate the correspondence of any variable to the kth group by the subscript k. For every neighborhood realization $\mathcal{Y}_k(\eta)$ belonging to the kth group, we have a fixed (conditional) probability for the pixel value to be 0 or 1 for a given set of parameters. Let $p_k = P(s = 1|\mathcal{Y}_k(\eta))$ and $q_k = 1 - p_k = P(s = 0|\mathcal{Y}_k(\eta))$. Then, using (2) and (3), we have $T_k = \ln(p_k/q_k)$, where $T_k = \underline{x}_k^T \underline{\omega}$. Hence, we obtain a system of K linear (and consistent) equations in terms of the GMRF parameters as

$$\alpha + (u + u')\beta_h + (v + v')\beta_c + (t + t')\beta_m$$
$$+ (w + w')\beta_r = d_k \ k = 1, \cdots, K, \tag{5}$$

where $d_k = \ln(p_k/q_k)$.

To estimate $\underline{\omega}$, the d_k's in (5) will be estimated first. This is done through histogramming, as follows. Suppose that there are a total of L_k pixels in a given coding $\mathcal{C}$ of $\mathcal{D}$ with neighborhood realizations from the kth group. From (1), the pixel values corresponding to these L_k locations become statistically independent when conditioned on the numerical realization of the set $\mathcal{D} \setminus \mathcal{C}$. Let $N_{1,k}$ be the number of 1 outcomes in these L_k observations. It is assumed that $\mathcal{Y}(\mathcal{D} \setminus \mathcal{C})$ is given; hence L_k is deterministic and $N_{1,k}$ is the outcome of a random quantity. Then, an estimate of d_k in (5) can be given as

$$\hat{d}_k = \ln\left(\frac{N_{1,k}}{L_k - N_{1,k}}\right) \text{ for } N_{1,k} \neq 0 \text{ and } N_{1,k} \neq L_k. \tag{6}$$

To justify this estimate, note that for large L_k, we have the asympthotic behavior

$$\frac{N_{1,k}}{L_k - N_{1,k}} \longrightarrow \frac{p_k}{q_k}. \tag{7}$$

Using (5) and (6), to estimate $\underline{\omega}$ we may try to solve the system of equations given as

$$\underline{x}_k^T \underline{\omega} = \hat{d}_k, \tag{8}$$

which may be put into matrix form as

$$X_r \underline{\omega} = \hat{\underline{d}} \tag{9}$$

with $\underline{x}_k^T$ and $\hat{d}_k$ being the corresponding rows of X_r and $\hat{\underline{d}}$, respectively. In general, it is not necessary to include the equations corresponding to every possible neighborhood realization to the system of equations in (9). The system in (9) will usually contain less than K equations. This is because for $N_{1,k} = 0$ or $N_{1,k} = L_k$,

(6) and (8) will be undefined and the corresponding image data will be discarded. This is one of the main weaknesses of the H method, as will be discussed in forthcoming sections. Furthermore, for some group of neighborhood realizations, L_k may be zero. Using (9), an estimate $\hat{\omega}_H$ of the parameter vector may be obtained as $\hat{\omega}_H = X_r^+ \hat{d}$, where X_r^+ is the pseudoinverse of the matrix X_r. If X_r is a full (column) rank matrix, then $X_r^+ = (X_r^T X_r)^{-1} X_r^T$. If X_r is not full rank, the resulting parameter estimates are generally highly corrupted.

For certain multilevel GMRF's, the above formulation can be simply modified by considering every pair of pixel values with a given neighborhood realization [7]. Also, the histogramming may, in practice, be done over the whole image instead of a single coding.

IV. RELATION OF THE H METHOD TO THE ML METHOD

The coding form of the ML estimate $\hat{\omega}_{ML}$ maximizes the probability $P(\mathcal{Y}(\mathcal{C}|\mathcal{Y}(\mathcal{D}\setminus\mathcal{C})))$ or, equaivalently, the *log-likelihood function* (LLF) [11]–[13]

$$\mathcal{L}(\mathcal{C}) = \sum_{i \in \mathcal{C}} \ln\left(P(y(i)|\mathcal{Y}(\eta_i))\right) \tag{10}$$

as a function of $\underline{\omega}$. The terms in (10) can be grouped under K terms, each of which corresponds to one of the K groups of neighborhood realizations as described in Section III. So we have

$$\mathcal{L}(\mathcal{C}) = \sum_{k=1}^{K} \mathcal{L}_k(\mathcal{C}), \tag{11}$$

where

$$\mathcal{L}_k(\mathcal{C}) = N_{1,k} \ln\left(P(y(i) = 1|\mathcal{Y}_k(\eta))\right)$$
$$+ (L_k - N_{1,k}) \ln\left(P(y(i) = 0|\mathcal{Y}_k(\eta))\right). \tag{12}$$

Although the discussion can be easily extended to multilevel GMRF's described in [7], we will consider the binary model given by (2) and (3), for which we have

$$\mathcal{L}_k(\mathcal{C}) = N_{1,k} \ln\left(\frac{e^{T_k}}{1+e^{T_k}}\right) + (L_k - N_{1,k}) \ln\left(\frac{1}{1+e^{T_k}}\right), \tag{13}$$

where $T_k = \underline{x}_k^T \underline{\omega}$. The gradient of $\mathcal{L}_k(\mathcal{C})$ denoted as $\nabla \mathcal{L}_k(\mathcal{C})$ is given by

$$\nabla \mathcal{L}_k(\mathcal{C}) = \frac{1}{1+e^{T_k}}\left(N_{1,k} - (L_k - N_{1,k})e^{T_k}\right)\underline{x}_k, \tag{14}$$

and $\nabla\nabla\mathcal{L}_k(\mathcal{C})$, the second derivative matrix of $\mathcal{L}_k(\mathcal{C})$, is given by

$$\nabla\nabla\mathcal{L}_k(\mathcal{C}) = \frac{\partial^2 \mathcal{L}_k(\mathcal{C})}{\partial\underline{\omega}\partial\underline{\omega}} = -\gamma_k \underline{x}_k \underline{x}_k^T, \text{ where } \gamma_k = \frac{e^{T_k} L_k}{(1+e^{T_k})^2}. \tag{15}$$

The value of γ_k is strictly positive for any $L_k > 0$; hence $\nabla\nabla\mathcal{L}_k(\mathcal{C})$ is negative semidefinite, with one negative eigenvalue $-\gamma_k$ corresponding to the eigenvector $\underline{x}_k$, and the remaining eigenvalues are zero. So $\mathcal{L}_k(\mathcal{C})$ is a convex cap function of the parameter vector $\underline{\omega}$. From (13), it is easy to see that $\mathcal{L}_k(\mathcal{C})$ has constant values on $(m-1)$-dimensional hyperplanes in the parameter space where m is the dimension of the parameter vector, $\underline{\omega}$ ($m = 5$ for the model (2), (3)). From (11), we also deduce that the LLF is a convex cap function. From (14), it is clear that $\mathcal{L}_k(\mathcal{C})$ will have a maximum value only if $N_{1,k} \neq 0$ and $N_{1,k} \neq L_k$. In any case, $\mathcal{L}_k(\mathcal{C})$ is upper bounded. In Fig. 2, a typical $\mathcal{L}_k(\mathcal{C})$ is plotted as a function of β_h and β_v.

If there exists a set of m groups of neighborhood realizations for which $N_{1,k} \neq 0$ and $N_{1,k} \neq L_k$, and if the corresponding coefficient vectors $\underline{x}_k$ are linearly independent, then the LLF will have a unique maximum (for which $\hat{\omega}_{ML}$ has a finite Euclidean norm). This fact is easily verified by noting from (11), (15) that, under the above conditions, the second derivative matrix of LLF will be negative definite (hence LLF will be strictly convex cap), and also the value of LLF will be decreasing in any direction in the parameter space as
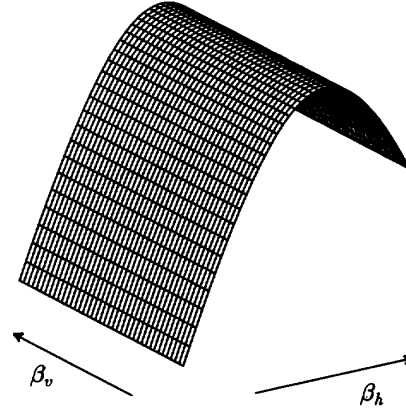

Fig. 2. A typical $\mathcal{L}_k(C)$ surface as a function of β_h and β_v($\beta_m = \beta_r = 0$) and assuming $N_{1,k} \neq 0$ and $N_{1,k} \neq L_k$.

the norm of $\underline{\omega}$ goes to infinity. However, the LLF may have a unique maximum point under much weaker conditions.

Using (5) and (14), it is easy to show that if the asymptotic behavior in (7) holds, then the hyperplanes over which $\mathcal{L}_k(\mathcal{C})$'s achieve their maxima will intersect at the point corresponding to the parameter vector of the GMRF under consideration. At this point, the LLF will achieve its global maximum. This suggests that an interesting approach to estimating $\underline{\omega}$ may be to search for an intersection for these hyperplanes. For a given $\mathcal{L}_k(\mathcal{C})$ for which $N_{1,k} \neq 0, L_k$, the hyperplane equation can be obtained by equating the gradient in (14) to zero, which yields the linear equation in $\underline{\omega}$

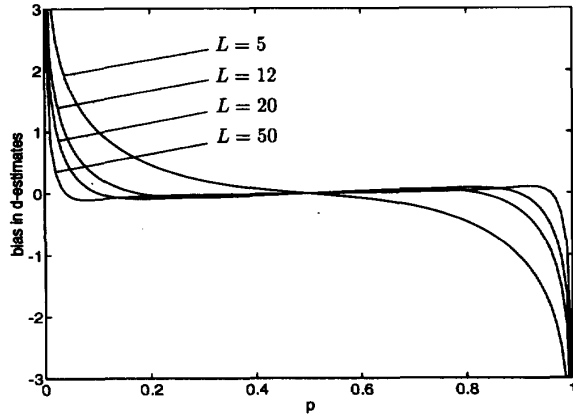$$T_k = \ln\left(\frac{N_{1,k}}{L_k - N_{1,k}}\right). \tag{16}$$

Note that the resulting system of linear equations is exactly the same as the system given by (8), and it may be solved by a least squares method. Hence we obtained the H method. In summary, ML method maximizes the LLF whereas H method tries to achieve this through maximizing individual terms $\mathcal{L}_k(\mathcal{C})$. This relationship of H and ML methods will still hold if the quantities $N_{1,k}$ and L_k are determined from the whole image region instead of a single coding.

Since the ML method requires optimization techniques, the H method seems computationally more feasible than the ML method, especially for applications involving parameter estimation over several subregions of a given image. On the other hand, since the H method discards the data corresponding to the case $N_{1,k} = 0$, and $N_{1,k} = L_k$, the parameter estimates may be highly corrupted especially in the case of parameter estimation over small subregions of a given image.

V. BIAS IN H METHOD

Any bias in the $\hat{d}$ vector in (9) is reflected in the $\hat{\omega}_H$ vector through the transformation $\hat{\omega}_H = X_r^+ \hat{d}$. In this section, we will formulate the bias for a given entry $\hat{d}_k$ of $\hat{d}$ vector for binary GMRF's. Since the analysis in this and the next section is done for an arbitrary but fixed group of neighborhood realizations, we will drop the subscript k, which refers to the kth group of neighborhood realizations. For a given L and p, and employing the independence property (1), N_1 can be shown to have binomial distribution

$$f_{N_1|L}(N_1|L) = \begin{cases} \binom{L}{N_1} p^{N_1}(1-p)^{L-N_1} & N_1 = 0, \cdots, L \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

Fig. 3.   Curves of bias $\mathcal{B}(\mathcal{L}, p)$ in $\hat{d}$ for various values of L.

However, in the H method, we omit the cases where $N_1 = 0$ or L. So the distribution is

$$g_{N_1|L}(N_1|L) = P(N_1|N_1 \neq 0 \text{ and } N_1 \neq L)$$
$$= \begin{cases} \frac{f_{N_1|L}(N_1|L)}{1 - p^L - (1-p)^L} & \text{for } N_1 = 1, \cdots, L-1 \\ 0 & \text{otherwise.} \end{cases} \quad (18)$$

Therefore, the expected value of $\hat{d}$ is given by

$$\mathcal{E}(L,p) = E\{\hat{D}(L, N_1)\} = \sum_{N_1=1}^{L-1} \ln\left(\frac{N_1}{L - N_1}\right) g_{N_1|L}(N_1|L), \quad (19)$$

where $E\{.\}$ is the statistical expectation operator and the notation $\hat{d} = \hat{D}(L, N_1)$ is introduced to emphasize the dependence of $\hat{d}$ on L and N_1. For any given L and p, the bias, $\mathcal{B}(L,p)$, in $\hat{d}$ is given by

$$\mathcal{B}(L,p) = \mathcal{E}(L,p) - d \quad \text{where} \quad d = \ln\left(\frac{p}{1-p}\right). \quad (20)$$

In Fig. 3, bias curves are plotted as a function of p and for several values of L. From the curves, it is seen that for large L the bias is smaller.

VI. REDUCTION OF BIAS IN THE H METHOD

As emphasized by the notation $\hat{d} = \hat{D}(L, N_1)$, $\hat{d}$ is a function of L and N_1, which may be interpreted as a look-up table indexed by L and $N_1 = 1, \cdots, (L-1)$. In this section, we will consider another estimate d^* of d as a function of L and N_1, i.e.,

$$d^* = D^*(L, N_1) \qquad N_1 = 0, \cdots, L. \quad (21)$$

The problem may be interpreted as the design of a new look-up table $D^*(L, N_1)$. The design criteria will be bias reduction, and we will include cases $N_1 = 0$ and $N_1 = L$. The bias $\mathcal{B}^*(L,p)$ in d^* is given by $\mathcal{B}^*(L,p) = \mathcal{E}^*(L,p) - d$, where

$$\mathcal{E}^*(L,p) = \sum_{N_1=0}^{L} D^*(L, N_1) f_{N_1|L}(N_1|L)$$
$$= \sum_{N_1=0}^{L} D^*(L, N_1) \binom{L}{N_1} p^{N_1} (1-p)^{L-N_1} \quad (22)$$

is the expected value of $D^*(L, N_1)$. Note that for given L, $\mathcal{E}^*(L,p)$ is a polynomial in p of order L. On the other hand, the variable d given in (20) can be represented only by an infinite order polynomial. Therefore, $\mathcal{B}^*(L,p)$ can not be made identically zero for all $p \in (0,1)$. So, instead of removing bias, we may try to choose $D^*(L, N_1)$

TABLE I
LOOK-UP TABLE EXAMPLES FOR $L = 5$

N_1	$\hat{D}(L, N_1)$	$D^*(L, N_1)$
0	-	-3.1523
1	-1.3863	-0.5751
2	-0.4055	-0.7615
3	0.4055	0.7615
4	1.3863	0.5751
5	-	3.1523

so as to approximate d by $\mathcal{E}^*(L,p)$ as a function of p for a given L. Therefore the problem reduces to approximating the variable d with a finite order polynomial. To accomplish this, we may force the bias to be zero for a finite number of values for p, i.e., $\mathcal{B}^*(L, p_j) = 0$ $j = 0, \cdots, J$ where $p_j \in (0, 1)$ are some chosen values for p. Hence, we obtain linear equations in terms of the to be determined variables $D^*(L, N_1), N_1 = 0, \cdots, L$ as

$$\sum_{N_1=0}^{L} D^*(L, N_1) f_{N_1|L}(N_1|L) = \ln\left(\frac{p_j}{1 - p_j}\right) \qquad j = 0, \cdots, J. \quad (23)$$

The above system of linear equations may be solved by a linear least squares method. After solving for the unknown $D^*(L, N_1)$ variables, one may put the results as a look-up table, and for any given L and N_1 one may use $D^*(L, N_1)$ as an estimate of d. For the estimation of $\underline{\omega}$, the system of equations in (8) may be rewritten as

$$\underline{x}_k^T \underline{\omega} = d_k^* \quad \text{where} \quad d_k^* = D^*(L_k, N_{1,k}). \quad (24)$$

The choice of p_j's in (53) determines the bias and mean-squared-error properties of the estimator. Since it is desirable to have $D^*(L, N_1) = -D^*(L, L - N_1)$ $n = 0, \cdots, L$, the p_j's must be chosen symmetrically around $p = 0.5$. In Table I, a sample look-up table for $L = 5$ is given both for the H method (i.e., $\hat{D}(L, N_1)$) and for the modification proposed in this section (i.e., $D^*(L, N_1)$). To compute $D^*(L, N_1)$, we have used

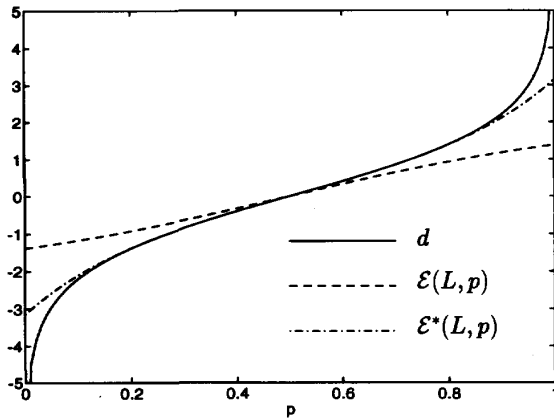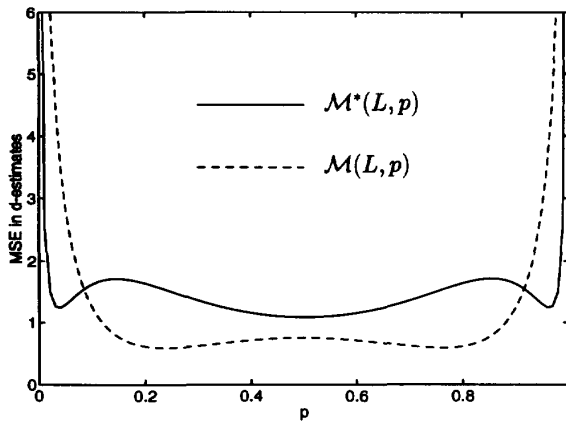$$p_j \in \{p \mid p = 0.5 \pm (0.05 + 0.01k) \; k = 0, \cdots, 15\}.$$

In Fig. 4, curves of d, $\mathcal{E}(L,p)$, and $\mathcal{E}^*(L,p)$ are plotted as a function of p for $L = 5$ (the look-up table in Table I is used). Note that the bias corresponds to the vertical difference between the d-curve and the $\mathcal{E}(L,p)$ or $\mathcal{E}^*(L,p)$ curves. At this point, it must be emphasized that in general the matrix X_r in (9) will be different in the H method and its modification proposed here (since we include the equations corresponding to $N_1 = 0, L_k$). Therefore, the effects of bias in $\hat{d}$ and d^* on the estimate of $\underline{\omega}$ will not be directly comparable, as in Fig. 4.

The mean squared errors (mse's) in $\hat{d}$ and d^* are given by

$$\mathcal{M}(L,p) = E\left\{(\hat{d} - d)^2\right\} \qquad \mathcal{M}^*(L,p) = E\left\{(d^* - d)^2\right\}, \quad (25)$$
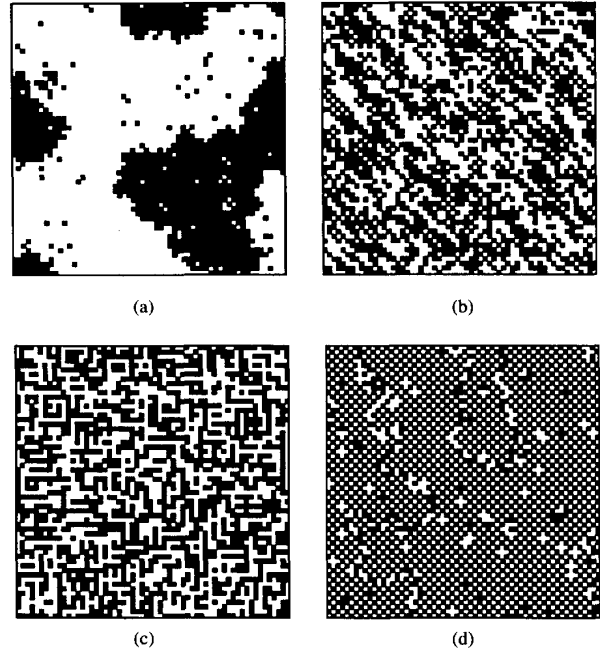
respectively. Both expectations are with respect to N_1. Note that in $\mathcal{M}(L,p)$, the distribution is given by (18), and for $\mathcal{M}^*(L,p)$ the distribution is given by (17). In Fig. 5, mse curves of d^* and $\hat{d}$ are plotted as a function of p for $L = 5$ (the look-up table in Table I is used). Again, it must be noted that the effects of the mse in $\hat{d}$ and d^* on the estimation of $\underline{\omega}$ will not be directly comparable from Fig. 5, because the X_r matrixes for the H method and its modification proposed here will be in general different.

In this section, we basically proposed a modification to the H method by using the bias reduction criterion. In this alternative method, we included the case of $N_1 = 0$ and $N_1 = L$ to the estimation process. Therefore, we have attacked two disadvantages of the H method simultaneously, namely the bias and the case of $N_1 = 0$ and $N_1 = L$.

Fig. 4. Curves of $d, \mathcal{E}(L,p)$, and $\mathcal{E}^*(L,p)$ for $L = 5$.

Fig. 5. Mse curves of d^* and $\hat{d}$ for $L = 5$.

Estimation of parameters over small subregions of a given image is usually required in many problems such as texture segmentation. In such circumstances, due to the limited amount of image data, it is very likely that the case of $N_1 = 0$ or $N_1 = L$ will be met. Since in the H method the corresponding data is discarded and no linear equation is obtained for the corresponding classes of neighborhood realizations, the system of linear equations in (9) usually contains a highly limited number of equations to give good parameter estimates. It is even very likely that the matrix X_r may not have full column rank, in which case the linear least-squares solution will not be unique (the transformation $\hat{\omega}_H = X_r^+ \hat{d}$ will give the minimum norm solution). This may result in highly corrupted parameter estimates. The number of rows of X_r may significantly increase by the inclusion of $N_1 = 0$ and $N_1 = L$ into the parameter estimation. For the above reasons, we included the case of $N_1 = 0$ and $N_1 = L$ in our look-up table design. Hence, it is more likely to have a full column rank X_r matrix.

Although this new estimator for d may introduce a larger mse for a range of p values, the combined effect of the increased number of rows in X_r and the averaging effect of the transformation $X_r^+ \hat{d}$ can be expected to reduce the extra mse on the parameter estimates. Therefore, the bias reduction approach proposed here seems to be a very reasonable criterion for the look-up table design. Also note that once the look-up tables are prepared, the computational complexities of the H-method and its modification will be almost the same.

Fig. 6. Sample GMRF textures with parameters: (a) $\alpha = -4$, $\beta_h = \beta_v = \beta_m = \beta_r = 1$. (b) $\alpha = -1.5$, $\beta_h = \beta_v = 0$, $\beta_m = 1.5$, $\beta_r = 0$. (c) $\alpha = 0$, $\beta_h = beta_v = 1$, $\beta_m = \beta_r = -1$. (d) $\alpha = 0$, $\beta_h = \beta_v = -1$, $\beta_m = \beta_r = 1$.

TABLE II
PARAMETER ESTIMATION RESULTS

Specified parameters		α	β_h	β_v	β_m	β_r		
		-4.00	1.00	1.00	1.00	1.00	T	T_s
H, 30×30 region	$\hat{\theta}$	-3.70	0.52	0.78	1.11	1.38	1000	87
	$\hat{\sigma}^2$	1.42	3.44	3.21	2.23	1.99		
Modified-H, 30×30 region	$\hat{\theta}$	-3.76	0.87	0.94	0.96	1.06	1000	711
	$\hat{\sigma}^2$	1.90	1.71	1.70	1.33	1.23		
H, 40×40 region	$\hat{\theta}$	-3.68	0.86	0.75	1.05	1.04	1000	628
	$\hat{\sigma}^2$	0.89	1.36	1.56	1.07	1.21		
Modified-H, 40×40 region	$\hat{\theta}$	-3.75	0.91	0.90	0.99	0.97	1000	959
	$\hat{\sigma}^2$	0.44	0.69	0.81	0.72	0.69		
H, 50×50 region	$\hat{\theta}$	-3.62	0.87	0.93	0.92	0.89	1000	951
	$\hat{\sigma}^2$	0.18	0.41	0.33	0.33	0.32		
Modified-H, 50×50 region	$\hat{\theta}$	-3.79	0.91	0.98	0.95	0.94	1000	994
	$\hat{\sigma}^2$	0.20	0.33	0.36	0.34	0.35		

VII. EXPERIMENTAL RESULTS

In this section, we will present experimental results corresponding to four different sets of specified parameters and various image region sizes. The test images have been synthetically generated by the Gibbs sampler algorithm using the model in (2) and (3) [5]. Typical sample textures are shown in Fig. 6 for each parameter set.

Experimental results are summarized in Tables II–V. The first column of each table indicates the method of parameter estimation and size of the image region (only one of the four codings corresponding to a second-order neighborhood structure has been used [11]). Tables II–V include the results for both the H method described in Section III and its modification (referred to as *modified-H* on the tables) based on the bias reduction criterion described in Section VI. The columns labeled as T indicate the number of times the corresponding experiment has been performed and T_s is the number of successful trials. By a successful trial we mean a performance of the experiment in which the matrix X_r comes out to be full rank (so that the least-squares solution of the resulting linear equations will be unique).

TABLE III
PARAMETER ESTIMATION RESULTS

Specified parameters		α	β_h	β_v	β_m	β_r		
		-1.50	0.00	0.00	1.50	0.00	T	T_s
H, 30×30 region	$\hat{\eta}$	-1.21	0.00	0.00	1.20	-0.01	1000	1000
	$\hat{\sigma}^2$	0.56	0.22	0.20	0.12	0.16		
Modified-H, 30×30 region	$\hat{\eta}$	-1.52	0.05	-0.01	1.48	0.00	1000	1000
	$\hat{\sigma}^2$	0.75	0.30	0.28	0.20	0.20		
H, 40×40 region	$\hat{\eta}$	-1.27	0.01	0.01	1.28	-0.01	1000	1000
	$\hat{\sigma}^2$	0.15	0.05	0.05	0.03	0.04		
Modified-H, 40×40 region	$\hat{\eta}$	-1.50	0.00	0.01	1.50	-0.01	1000	1000
	$\hat{\sigma}^2$	0.24	0.08	0.08	0.07	0.06		
H, 50×50 region	$\hat{\eta}$	-1.37	0.00	0.00	1.36	0.01	1000	1000
	$\hat{\sigma}^2$	0.08	0.02	0.02	0.02	0.02		
Modified-H, 50×50 region	$\hat{\eta}$	-1.51	0.00	0.00	1.51	-0.01	1000	1000
	$\hat{\sigma}^2$	0.12	0.04	0.04	0.03	0.03		

TABLE IV
PARAMETER ESTIMATION RESULTS

Specified parameters		α	β_h	β_v	β_m	β_r		
		0.00	1.00	1.00	-1.00	-1.00	T	T_s
H, 20×20 region	$\hat{\eta}$	-0.13	1.00	0.88	-0.73	-1.00	1000	47
	$\hat{\sigma}^2$	6.36	2.31	1.29	1.10	1.02		
Modified-H, 20×20 region	$\hat{\eta}$	0.18	1.09	0.93	-0.91	-1.35	1000	107
	$\hat{\sigma}^2$	14.55	2.49	3.03	2.91	4.12		
H, 30×30 region	$\hat{\eta}$	-0.03	0.80	0.82	-0.77	-0.80	1000	1000
	$\hat{\sigma}^2$	0.65	0.17	0.16	0.15	0.14		
Modified-H, 30×30 region	$\hat{\eta}$	0.02	0.97	0.99	-0.96	-1.00	1000	1000
	$\hat{\sigma}^2$	0.91	0.24	0.24	0.20	0.21		
H, 40×40 region	$\hat{\eta}$	0.04	0.81	0.82	-0.82	-0.84	1000	1000
	$\hat{\sigma}^2$	0.17	0.05	0.05	0.04	0.04		
Modified-H, 40×40 region	$\hat{\eta}$	0.02	0.98	0.98	-0.98	-0.99	1000	1000
	$\hat{\sigma}^2$	0.26	0.08	0.07	0.06	0.06		

TABLE V
PARAMETER ESTIMATION RESULTS

Specified parameters		α	β_h	β_v	β_m	β_r		
		0.00	-1.00	-1.00	1.00	1.00	T	T_s
H, 20×20 region	$\hat{\eta}$	-	-	-	-	-	1000	0
	$\hat{\sigma}^2$	-	-	-	-	-		
Modified-H, 20×20 region	$\hat{\eta}$	0.43	-0.53	-1.23	0.94	0.46	1000	21
	$\hat{\sigma}^2$	16.98	2.34	3.91	1.88	1.17		
H, 30×30 region	$\hat{\eta}$	-0.32	-0.86	-0.78	0.90	1.18	1000	78
	$\hat{\sigma}^2$	16.19	4.28	2.91	2.92	2.11		
Modified-H, 30×30 region	$\hat{\eta}$	0.13	-1.08	-0.87	0.94	0.90	1000	706
	$\hat{\sigma}^2$	· 9.10	1.90	1.67	1.32	1.18		
H, 40×40 region	$\hat{\eta}$	-0.46	-0.84	-0.77	1.07	1.04	1000	637
	$\hat{\sigma}^2$	7.80	1.44	1.51	1.15	1.12		
Modified-H, 40×40 region	$\hat{\eta}$	-0.02	-0.96	-0.91	0.97	0.94	1000	959
	$\hat{\sigma}^2$	4.52	0.92	0.88	0.73	0.75		

Note that in the modified-H method, the success rate is always at least as large as the success rate of the H method. This is because the modified-H method uses the data for which $N_1 = 0$ or L for a set of L values, as will be described in this section. The sample means $\hat{\eta}$ and variances $\hat{\sigma}^2$ of each parameter is computed over the successful trials only. More specifically,

$$\hat{\eta} = \frac{1}{T_s}\sum_{i=1}^{T_s}\xi_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{T_s}\sum_{i=1}^{T_s}(\xi_i - \hat{\eta})^2,$$

where ξ_i denotes the value of a parameter $\xi \in \{\alpha, \beta_h, \beta_v, \beta_m, \beta_r\}$ for the ith successful trial.

For both the H and modified-H methods, we have discarded the data corresponding to $L < 5$. In the modified-H method, look-up tables based on the bias reduction criterion have been generated only for $L = 5, 6, \cdots, 11$. For $L > 11$, the estimator of d given in (6) is used for both methods. This is because an important portion of the discarded data in the H method usually corresponds to

$L = 5, 6, \cdots, 11$ for small image region sizes. Furthermore, the bias is not very significant for $L > 11$.

For generating the look-up tables, the p_j values in (23) have been taken as

$$p_j \in \{p \mid p = 0.5\pm(0.05 + 0.01k)\ k = 0, \cdots, 15\}$$

$$\text{for } L = 5, 6, \cdots, 10,$$

$$\text{and } p_j \in \{p \mid p = 0.5\pm 0.01k\ k = 0, \cdots, 25\}$$

$$\text{for } L = 11.$$

Although no specific strategy has been followed in choosing the p_j values as given above, we have obtained good results in many experiments for small image region sizes.

The results summarized in Tables II–V indicate that the modified-H method usually yields a lower bias and also that the success rate is higher than the H method. The sample variances indicate that, for many applications, the H and modified-H methods have a reasonable variance. It must be noted at this point that once the look-up tables are prepared, the computational complexity of the method proposed in this correspondence and the H method are almost the same.

VIII. CONCLUSION

An advantage of the H method over the ML method is in its computational simplicity. The ML method requires the maximization of a log-likelihood function, whereas the H method involves simple histogramming, a look-up table operation, and computation of the pseudoinverse of a matrix with generally reasonable dimensions. For its computational simplicity, the H method is well suited to problems involving parameter estimation over several subregions of a given image. On the other hand, the omission of data (for $N_{1,k} = 0, L_k$) in the H method may result in very poor performance over small image regions. Even more, the X_r matrix may not have full column rank. However, in problems such as image segmentation it is usually required to estimate the GMRF parameters over several small subregions of the given image. Therefore it is very important to have an estimator that is both computationally simple and capable of yielding sufficiently good parameter estimates even in the case of a small amount of image data.

In this correspondence, we proposed a modification to the H method to improve its performance, especially in the case of a small number of observations while preserving its computational advantage over the ML method. This modification is computationally almost as easy as the H method. Furthermore, it does not suffer from the data omission and hence it is more likely that X_r matrix will have full column rank. The extension of the method to multilevel GMRF's and the methodology for choosing p_j values in (23) may be topics of future research.

REFERENCES

[1] R. L. Kashyap and A. Khotanzad, "A model-based method for rotation invariant texture classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 4, pp. 472–481, July 1986.
[2] K. Deguchi and I. Morishita, "Texture characterization and texture-based image partitioning using two-dimensional linear estimation techniques," *IEEE Trans. Comput.*, vol. C-27, no. 8, pp. 739–745, Aug. 1978.
[3] R. Chellappa and R. L. Kashyap, "Digital image restoration using spatial interaction models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, no. 3, pp. 461–472, June 1982.
[4] F. R. Hansen and H. Elliott, "Image segmentation using simple Markov field models," *Comput. Graphics Image Processing*, vol. 20, pp. 101–132, 1982.

[5] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[6] H. Derin and W. S. Cole, "Segmentation of textured images using Gibbs random fields," *Comput. Vision, Graphics, Image Processing*, vol. 35, pp. 72–98, 1986.

[7] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, no. 1, pp. 39–55, Jan. 1987.

[8] C. S. Won and H. Derin, "Unsupervised segmentation of noisy and textured images using Markov random fields," *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 4, pp. 308–328, July 1992.

[9] F. S. Cohen and Z. Fan, "Maximum likelihood unsupervised textured image segmentation," *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 3, pp. 289–251, May 1992.

[10] N. Ahuja and A. Rosenfeld, "Mosaic models for textures," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, no. 1, pp. 1–11, Jan. 1981.

[11] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, no. 1, pp. 25–39, Jan. 1983.

[12] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc., Series B*, vol. 36, pp. 192–236, 1974.

[13] ——, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc., Series B*, vol. 48, pp. 259–302, 1986.

[14] C. O. Acuna, "Texture modeling using Gibbs distributions," *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 3, pp. 210–222, May 1992.

Contribution to the Determination of Vanishing Points Using Hough Transform

Evelyne Lutton, Henri Maître, and Jaime Lopez-Krahe

Abstract—We propose a method to locate three vanishing points on an image, corresponding to three orthogonal directions of the scene. This method is based on two cascaded Hough transforms. We show that, even in the case of synthetic images of high quality, a naive approach may fail, essentially because of the errors due to the limitation of the image size. We take into account these errors as well as errors due to detection inaccuracy of the image segments, and provide a method efficient, even in the case of real complex scenes.

Index Terms—Bias and errors of the Hough transform, Hough transform, orthogonal directions detection, vanishing points detection.

I. INTRODUCTION

In many tasks of artificial vision, an accurate location of vanishing points is a first step toward three-dimensional (3D) interpretation. Vanishing points are defined in the image plane as those points where the images of all 3D scene lines, parallel to some space direction, converge. To one 3D space direction is attached one vanishing point on the image plane and conversely.

Detection of vanishing points, which is of little help in natural outdoor scenes, becomes of prime importance in the man-made environment where regular block looking structures or parallel alignments (streets, pavements, railroad) abound. We have the Italian Quattrocento to thank for the deep comprehension of the formation
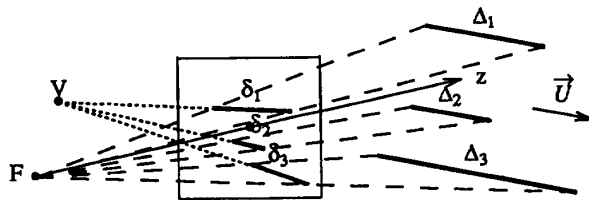
Fig. 1. Δ_1, Δ_2, and Δ_3 are lines of the 3D space parallel to the direction $\vec{U}$. δ_1, δ_2, and δ_3 their images on the image plane. δ_1, δ_2, and δ_3 converge to the vanishing point associated to $\vec{U}$.

of perspective images,[1] a comprehension that has been a constant preoccupation of the theoricians in aesthetics up to the 20th century.[2] Traces of rigorous mathematical bases are found mainly at the corner of the 18th–19th centuries.[3,4] Although the geometrical construction of vanishing points may become very complex when no hypothesis is made on the vision system, in the case of a perfect conic projection (pinhole cameras) it may be solved easily, since the image of a straight line remains a straight line. We will stay in the assumption of conic projection throughout this paper.

Most of the existing methods to detect automatically these vanishing points stand on the use of the Hough transform, explicitly or not [16], [20], [21]. Hough transform is a global technique for detecting parametrical structures in images [5], [7], [17]. Some primitives are detected in the image, and then mapped into a parameter space; underlying structures are detected by searching for clusters in this parameter space. Two methods can be distinguished to fill the accumulators of Hough space [17].

- The "one-to-many" (1-to-m) transform, used most of the time, where for each feature point in the picture plane, several accumulator cells are incremented.
- The "many-to-one" (m-to-1), where we make use of several feature points in the image plane to increment exactly one accumulator cell.

For the application of vanishing points detection, the primitives are line segments. The vanishing points are thus characterized as those points where most of the supporting lines of these segments intersect (Fig. 1). Most of the time, these points are located far away from the image limits and even can be at infinity (for frontal lines).

So, the most important problem of the detection of vanishing points with the help of the Hough transform is the choice of the Hough space parameterization. Two main orientations have appeared in the literature, following the 1-to- m or the m-to-1 transforms; they are chronologically:

The 1-to-m approach with:

- Kender [9] in 1979 who uses, directly on the image plane, circles passing through the origin, and proposes either a search in a tridimensional space or two successive transforms;
- Ballard and Brown [2] in 1982, who propose a $(k/r, \theta)$ parameterization, with k constant, for the image primitives, which permits restriction of parameter space;

[1] L. B. Alberti, "De pictura" (manuscript 1435), printed in Basl (Swiss), 1540.

[2] E. Panovsky. "Die Perspektive als symbolische Form," Berlin, 1927.

[3] G. Monge, "Géométrie descriptive," *Leçons données aux Ecoles Normales de l'an III de la République*, Paris, 1798.

[4] J. V. Poncelet, "Traité des propriétés projectives des figures," Paris, 1822.