*Indian J. pure appl. Math.*, **25** (1&2); 239-245, January & February 1994

# TOWARDS SEMI-PARAMETRIC IMAGE-PROCESSING

U. V. NAIK-NIMBALKAR AND M. B. RAJARSHI

*Department of Statistics, University of Poona, Pune* 411 007

*(Accepted* 10 *May* 1993*)*

Statistical image processing has been largely confined to model-based approach. Here, we discuss a semi-parametric approach based on the theory of estimating functions. The true image is seen through colour intensities at pixels. It is assumed that these intensities follow a random process such as a Markov Random Field. The true colouring is unknown, but associated with each pixel, there is a record conveying imperfect information about the colour, according to a statistical model. We do not assume that the joint distribution of the colour-intensities and records is completely specified. In stead, the first two conditional moments are specified. An illustrative example is given. Our method assumes that the intensity is a continuous variable. We also indicate how to deal with the situation, wherein the true intensity can take only finitely many values.

## 1. INTRODUCTION

In most of the statistical image processing, a parametric model is assumed both for the unknown image and for its observed record or degraded image. In fact, it is customary to assume that the true image $x$ is a realisation of a Gaussian random field and the observed image $y$ given $x$, has a Gaussian distribution. For a good account of statistical image processing, we refer to Besag[1] and Karr[2], who assume that the joint distribution of the records and the image is completely known. In this note, we use the Estimating Function(EF) methodology for estimation of the true image. In this set-up only the first two conditional moments of the intensities (true image) and observables need to be specified. Our results are based on applications of estimating functions to the filtering and the state-estimation as reported in Naik-Nimbalkar and Rajarshi[3]. In this approach, a system of simultaneous equations is solved to estimate the true image (or scores associated with it) and optimality of the procedure is defined in terms of the optimality of the equation itself. The approach of estimating functions in the statistical methods was initiated by Godambe[4]. The optimality criteria for the estimating function was suggested by Godambe[4] and Durbin[5] and has been adopted by a number of authors. For a recent account of this theory, its comparison with the Likelihood and the (weighted) Least Squares approach and its similarly with the Quasi-likelihood approach, we refer to Godambe[6]. Our note uses a modification of a result from Godambe and Thompson[7].

In Section 2, we outline the parametric approach due to Besag[1]. The semi-parametric or distribution-free solution based on estimating functions and its links with the parametric solution are discussed in Section 3. The last section discusses problems associated with a digital image, i.e., when only a finitely many

scores are associated with each pixel.

## 2. Parametric or Model-Based Approach

The true image is modelled as a random field $X$, i.e., a collection of random variables on a finite, two dimensional lattice $S = \{(i, i) \mid i = 1, 2, ..., n; j = 1, 2, .., n\}$. Let $X = \{X(i, j) \mid (i, j) \in S\}$, where each $X(i, j)$ us a real number. Following Besag[1], section 4, we interpret $X(i, j)$ as a grey-score or intensity, and, we deal with 'continuous' colour. Ley $Y = \{ Y(i, j), i = 1, 2, .., n; j = 1, 2, ..., n\}$ denote the observable random record. Realized values are denoted by corresponding lower case letters such as $x$ and $y$.

**Assumptions** :

1. For every realised $x$, the random variables $Y(i, j)$, $i = 1, 2, .., n$; $j = 1, 2, .., n$ are conditionally independent. Each $Y(i, j)$ has a known conditional density function $f\{y(i, j) \mid x(i, j)\}$, which depends only on $x(i, j)$.

2. The true colouring $x$ is a realization of a locally dependent Markov random filed (MRF) with specified probability density function $p(x)$. More precisely,

$$p(x(i, j) \mid x(r, s), r, s = 1, 2, .., n; (r, s) \neq (i, j))$$
$$= p(x(i, j) \mid x(r, s), (r, s) \in A(i, j)) \quad .. (2.1)$$

where $A(i, j)$ is a subset of $\{(r, s) \mid r, s = 1, 2, .., n, (r, s) \neq (i, j)\}$.

Then, members of the set $A(i, j)$ are called neighbours of the pixel $(i, j)$.

For example, one such model is defined by taking

$$A(i, j) = \{ (i, j -1), (i, j + 1), (i - 1, j), (i + 1, j) \}.$$

To estimate the true image $x$, one proceeds as follows. By Baye's theorem $p(x \mid y) = f(y \mid x)\, p(x)$. The estimate $\hat{x}$ of $x$ is the one which maximizes $p(x \mid y)$ and is known as Maximum a posteriori (MAP) estimate of $x$. Besag[1] discusses a number of algorithms to obtain the posterior distribution $p(x \mid y)$.

## 3. Semi-Parametric Solution Based on Estimating Functions

We now weaken the Assumptions 1 and 2 of Section 2, by requiring that only the first two conditional moments in each case are specified. More precisely, we specify

$$\mu(i, j) = E\,[Y(i, j) \mid x(i, j)] \text{ and } \sigma^2(i, j) = \text{var}\,[y(i, j) \mid x(i, j)], \text{ and}$$

$$E[X(i, j) \mid x(r, s), r, s) = 1, 2, .., n; (r, s) \neq (i, j) ]$$
$$= E[X(i, j) \mid x(r, s), (r, s) \in A(i, j) ] = M(i, j) \text{ (say)}$$

$$\text{Var } [X(i, j) \mid x(r, s), r, s = 1, 2, .., n; (r, s) \neq (i, j) ]$$
$$= \text{Var } [X(i, j) \mid x(r, s), (r, s) \in A(i, j) ] = V(i, j) \text{ (say)} \quad ... (3.1)$$

For convenience of illustration, the set $A(i, j)$ is taken to be $\{ (i, j-1), (i, j+1), (i-1, j), (i-1, j), (i+1, j) \}$ and the neighbours of the boundary points are assumed to be appropriately defined.

Based on the above assumptions, we derive a class of estimating functions. A brief introduction to the methodology of estimating functions is given below with more details in the Appendix.

Let $\mathcal{Y} = \{y\}$ be a sample space, $\mathcal{F}$, a $\sigma$-field and $\mathcal{P} = \{P\}$, an appropriate class of probability distributions on $(\mathcal{Y}, \mathcal{F})$. Let $\theta$ be a $p$-dimensional parameter and $\Theta = \{\theta\}$ denote the parameter space. A function $g : \mathcal{Y} \times \Theta \rightarrow R^p$ is said to be an unbiased estimating function, if $E_p [g(y, \theta)] = 0$, for every $P \in \mathcal{P}$. Usually, a class of unbiased estimating functions is obtained depending upon the model, and the 'optimal' function $g^*$ within this class is derived, see the Appendix. A solution of $g^* = 0$ estimates $\theta$.

For applications to the image processing problem, we note that the true image $x$, which is itself a realization of a random field, is to be estimated. Thus, $\theta$ in the above discussion coincides with $x$ and the situation is methodologically parallel to Ferreira[8]. We follow Ferreira[8] and interpret the expectation in the definition of an unbiased EF, as expectation with respect to $P$, the joint distribution of the random field $x$ and the observable $y$.

Let $h_1(i, j) = y(i, j) - \mu(i, j)$ and $h_2(i, j) = x(i, j) - M(i, j)$ for all $(i, j)$. Notice that $\mu(i, j)$ involves only $x(i, j)$, whereas $M(i, j)$ involves $\{x(r, s) \mid (r, s) \in A(i, j)\}$. To describe our class of estimating functions $\mathcal{G}$, we proceed as follows. Following Besag[9], we introduce the coding system $S_1$ and $S_2$, which is a partition of $S$ such that if $(i, j) \in S_k$, then $A(i, j) \subset S_l$, $l \neq k$ (cf. diagram 1). Consider the class of estimating functions $\mathcal{G} = \{ g(i, j) \mid (i, j) \in S \}$, where

$$g(i, j) = \sum_{(r, s) \in S} a_{ij}^{(1)}(r, s)\, h_1(r, s) + \sum_{(r, s) \in S_1} a_{ij}^{(2)}(r, s)\, h_2(r, s), \quad \ldots (3.2)$$

where $a_{ij}^{(1)}(r, s)$ is a function of $x$ and $a_{ij}^{(2)}(r, s)$ is a function of and $x(r', s')$, $(r', s') \in A(r, s)$. The coding $S_1$ and $S_2$ and the assumptions 1 and 2 ensure that the collection of functions $\{h_1(i, j), (i, j) \in S, h_2(i, j), (i, j) \in S_1\}$ satisfies the orthogonally condition of Theorem 2.1 of Godambe and Thompson[7]. It follows that the optimal EF within $\mathcal{G}$ is given by

$$\begin{aligned}
\tilde{a}_{ij}^{(1)}(r, s) &= \partial \mu(i, j)/\partial x(i, j)/\sigma^2(i, j), && \text{if } (r, s) = (i, j).\\
&= 0 && \text{otherwise;}\\
\tilde{a}_{ij}^{(2)}(r, s) &= -\{ \partial M(r, s)/\partial x(i, j) \}/V(r, s), && \text{if } (i, j) \in A(r, s),\\
&= 1/V(i, j) && \text{if } (r, s) = (i, j),\\
&= 0 && \text{otherwise.} \quad \ldots (3.3)
\end{aligned}$$

The posterior score function $\partial \log p(x \mid y)/\partial x(i,j)$, $(i,j) \in S$, if the distribution forms are known, is optimal within the class of unbiased EFs, under certain regularly conditions, cf. Ferreira[8] and Ghosh[10]. The parametric solution of Section 2 exactly corresponds to this and hence the MAP procedure as suggested in Besag[1] and Karr[2], is justified within the framework of EFs. It may be pointed out that under certain regularity conditions, the EF given by (3.3) is the projection of the posterior score function on the class $\mathcal{G}$, viewed as a subspace of the Hilbert space of square integrable (with respect to $P$) functions, cf. Naik-Nimbalkar and Rajarshi[3], Appendix. Thus, the optimal EF within a class $\mathcal{G}$ coincides with the posterior score function, if the latter is in the class $\mathcal{G}$.

Another class of estimating functions can be obtained by taking $(r, s)$ in $S_2$ for the second term of (3.2) and an optimal estimating function can be derived in a similar manner. Thus, corresponding to the two codings $S_1$ and $S_2$, we have two estimates of the true image $x$. As in spatial analysis, these two estimates can be combined, for example, by taking their mean.

According to the theory of EFs, one can optimally combine two EFs $g_1$ and $g_2$ provided the matrix $E[g_1\ g_2^{TRA}]$ and the matrices of expectations of partial derivatives of $g_1$ and $g_2$ with respect to $x(i, j)$, are known, cf. Naik-Nimbalkar, Lele and Godambe[11]. In fact, under the assumption of stationarity of the MRF $x$, the optimal EF turns out to be $g_1 + g_2$.

*Example* 3.1 — In image processing, the error variability in recording an image at a pixel $(i, j)$ may be related with the intensity of the true image at $(i, j)$. Low intensity image may be recorded with a large observational error. Thus, it may be reasonable to assume that $\sigma^2(i,j) = \mathrm{Var}[y(i,j) \mid x(i,j)]$ is a function of $x(i, j)$. This fact can be easily accomodated within our set-up. Consider the following with the notation as given above.

$$\mu(i,j) = x(i,j), \quad \sigma^2(i,j) = \eta/[x(i,j) + \delta];$$

$$M(i,j) = \theta\, z(i,j), \quad V(i,j) = \sigma^2,$$

where $z(i, j) = x(i, j-1) + x(i, j+1) + x(i+1, j) + x(i-1, i)$ and $\{\theta, \sigma^2, \eta, \delta\}$ are known para,eters. The coding system $S_1$ and $S_2$ with $n = 6$ is described below, where $X$ denotes a pixel in $S_1$ and 0 denotes a pixel in $S_2$.

| | | | | | |
|---|---|---|---|---|---|
| x | 0 | x | 0 | x | 0 |
| 0 | x | 0 | x | 0 | x |
| x | 0 | x | 0 | x | 0 |
| 0 | x | 0 | x | 0 | x |
| x | 0 | x | 0 | x | 0 |
| 0 | x | 0 | x | 0 | x |

**DIAGRAM 1**

Let us consider the coding $S_1$. Ignoring the pixels on the boundary in the second term of (3.2), the component $(i, j)$ of the optimal EF (3.3) is given by

$$-[y(i, j) - x(i, j)]\,[x(i, j) + \delta]/\eta + [x(i, j) - \theta z(i, j)]/\sigma^2 \quad \text{if } (i, j) \in S_1,$$

and

$$-[y(i, j) - x(i, j)]\,[x(i, j) + \delta]/\eta +$$

$$-\theta \sum_{(r, s) \in B} [x(r, s) - \theta z(r, s)]/\sigma^2 \quad \text{if } (i, j) \in S_2, \qquad \ldots (3.4)$$

where $B = \{ (r, s) \mid (i, j) \in A(r, s) \}$.

*Example* 3.2 — The following model is based on Karr[2] (Section 5). The observed image, indexed by pixels, has the form

$$y(i, j) = \phi(H(x)(i, j)) + \varepsilon(i, j), \qquad \ldots (3.5)$$

where (1) $H$ is a known blurring function, which operates on the entire undegraded pixel process $x$, (2) $\phi$ is a known, possibly non-linear transformation, applied pixel-by-pixel, (3) $\varepsilon(i, j)$ is a random noise with mean 0, uncorrelated with the noise at any other pixel and (4) the collection $\{\varepsilon(i, j)\}$ is statistically independent of $x$.

Though the function $H$ is a function of the entire true scene $x$, quite often, in image analysis, it is assumed to be "local" in nature. For example,

$$H(x)(i, j) = \gamma\, x(i, j) + (1 - \gamma) \sum_{(r, s) \in A(i, j)} x(r, s) / \mid A(i, j) \mid,$$

where $\mid S \mid$ denotes the number of elements in the set $S$. The assumption 2 of Section 2 continues to hold. The optimal estimating equation can be easily obtained as in Example 3.1 above. In this case, $\mu(i, j)$ depends on $x(i, j)$ as well as the intensities at the neighbours of $(i, j)$.

*Remark* 3.1 : Obtaining solution of equations based on EFs discussed above is quite involved. A recurssive approach, as discussed in Section 4.2.1 of Besag[1] can be adopted. Besag[1] suggests application of a parallel processing system to solve such problems.

*Remark* 3.2 : Throughout, we have assumed that the statistical parameters such as $\sigma, \eta, \delta, \theta$ etc. are known. This assumption is often not true in practice. It is often made to simplify the discussion. The following algorithm appears to be feasible. Let $\eta$ be the vector of unknown statistical parameters.

1. For a fixed value of $\eta$, estimate $x$ as described above. Let this be denoted by $x(\eta)$.
2. Obtain $W(\eta) = \Sigma\, [y(i, j) - x(\eta, i, j)]^2/\sigma^2\,(i, j)$.
3. Estimate $\eta$ by arg-min $W(\eta)$.

## 4. DISCRETE COLOURS

So far, we assumed that the colour in an image can vary continuously. However, in some situations, it is possible that the colour is only classified into $c + 1$ categories, see Section 4.1 of Besag[1]. Our approach in the last two sections assumed that the estimating functions involved are differentiable in $x(i, j)$'s and thus needs to be modified. If there is a natural ordering of colours, say from black to white, following Besag[1], we call them as grey levels. As Besag[1] points out, a typical image processing equipment has 256 levels and in such a case, one may continue to model the true scene as if the colour is continuous.

When such an approach is not feasible and the true pixel outcomes can be 0, 1, 2, ..., $c$, a possible solution based on the EF approach can be as follows. The EFs $h_1$ and $h_2$ are still valid but the $x(i, j)$'s do not vary in an open set. Following Lindsay and Roeder[12], if there is a natural extension of $h_1$ and $h_2$ to non-integer values of $x(i, j)$'s, then one can continue to differentiate $h_1$ and $h_2$ with respect to $x(i, j)$'s. Integers nearest to the so-obtained solutions are then taken as the estimates of the true image categories.

## REFERENCES


1. J. Besag, *J. Roy. Statist. Soc. B*, **36** (1986)259-302.
2. A. F. Karr, in *Satistical Inference in Stochastic Processes*, (ed. N. V. Prabhu & I. V. Basawa), Marcel Dekker, (1991) p.1-33.
3. U. V. Naik-Nimbalkar & M. B. Rajarshi, in *Proc. of Symp. in Honour of Professor V. P. Godambe*, University of Waterloo, Canada, (1992).
4. V. P. Godambe, *Ann. Math. Satist.* **31** (1960)1208-1212.
5. J. Durbin, *J. Roy. Statist. Soc. B*, **22**(1960)139-153.
6. V. P. Godamber (ed.), *Estimating Functions*, Oxford Univ. press (1991).
7. V. P. Godambe & M. E. Thompson, *J. Statist. Planning and Inference*, **22** (1989)137-152.
8. P. E. Ferreria, *Biometrika*, **69** (1982)192-236.
10. M. Ghosh, in *C. G. Khatri memorial volume of the Gujrat Statistical Review*, (1991).
11. U. V. Naik-Nimbalkar, S. Lele and V. P. Godambe, *Technical Report, Department of Statistics, Uni. Poona*, 89-1 (1989).
12. B. G. Lindsy and K. Roeder, *J. Amer. Statist. Assoc.* **82** (1987)758-764


## APPENDIX

Let the estimating function $g$ as described in Section 3 be $g = (g^{(1)}, g^{(2)}, \ldots, g^{(p)})^T$, where $T$ denotes the transpose of a matrix. Assume that $g^{(r)}$ is differentiable with respect to $\theta_j$ for all $r$ and $j$. Let the $p \times p$ matrices $\mathbf{J}$ and $\mathbf{H}$ be defined by

$$\mathbf{J} = E_p(gg^T); \quad \mathbf{H} = ((E_p(\partial g^{(r)}/\partial\theta_j))). \qquad \ldots (A.1)$$

Let $\mathbf{J}^*$ and $\mathbf{H}^*$ denote the matrices corresponding to an estimating function $g^*$. Then, we have :

*Definition* A.1. — We say $g^*$ is an optimal estimating function within $\mathcal{G} = \{g\}$, class of unbiased estimating functions. if

$$\mathbf{J} - \mathbf{H}\,(\mathbf{H}^*)^+\,\mathbf{J}^*\,(\mathbf{H}^{*\mathbf{T}})^+\,\mathbf{H}^T \qquad \ldots (A.2)$$

is positive semi-definite for every $g \in \mathcal{G}$. In (A.2), $A^+$ denotes the (unique) Moore-Penrose generalized inverse of the matrix $A$.

The above multi-parameter optimally criterion has been adopted by many authors, see Godambe and Thompson[7] for additional references. In our applications, the true scene $x$ is random and the definition (A.2) can be modified by interpreting the expectation $E$ as the expectation with respect to the joint distribution of $x$ and $y$, cf. Ferreira[8] and Ghosh[10]. The theorem 2.1 of Godambe and Thompson[7] is stated here for convenience.

*Theorem.* Let $\{\mathcal{F}_j\}$ be a collection of sub-σ-fields of $\mathcal{F}$. Let $h_j$ be a function on $\mathcal{Y} \times \Theta$, such that $E(h_j \mid \mathcal{F}_j) = 0$, $j = 1, 2, .., m$. Let $\mathbf{g}$ be an estimating function with the $r$-th component defined by

$$g^{(r)} = \sum_{j=1}^{m} h_j\, a_j^{(r)} \qquad \ldots (A.3)$$

$a_j^{(r)}$ being some real functions of $(y, \theta)$ which is $\mathcal{F}_j$-measurable. Let $\mathbf{g}^*$ be defined by

$$g^{(r)*} = \sum_{j=1}^{m} a_j^{(r)*} h_j\,, \qquad \ldots (A.4)$$

where

$$a_j^{(r)*} = \frac{E(\,(\partial h_j/\partial \theta_r) \mid \mathcal{F}_j)}{E\,(h_j^2 \mid \mathcal{F}_j)} \qquad \ldots (A.5)$$

Further, let $h_{jr}^* = h_j\, a_j^{(r)*}$. Assume that

$$E\,(h_{j,r}^*\, h_{j',r'}^*, \mid \mathcal{F}_i) = 0 \quad j \neq j'\,, \quad j, j' = 1, 2, .., m;\ r, r' = 1, 2, .., p. \qquad \ldots (A.6)$$

Then, $g^*$ is optimal in $\mathcal{G} = \{g\}$.

*Remark :* The condition (A.6) is known as mutual orthogoanlity of $h_j$'s.

The above theorem can be modified to the situation wherein $\theta$'s are realised values of random variables. For applications made in Section 3, the σ-field corresponding to $h_1\,(i, j)$'s is the σ-field generated by $\{x(r, s) \mid (r, s) \in A(i, j)\}$. The orthogonality of these functions is a consequence of the assumptions made, the coding and the properties of the conditional expectation.