# On the Detection of Edges in Vector Images

### Petar M. Djurić and Jong-Kae Fwu

Abstract— A novel method for edge detection in vector images is proposed that does not require any prior knowledge of the imaged scenes. In the derivation, it is assumed that the observed vector images are realizations of spatially quasistationary processes, and that the vector observations are generated by parametric probability distribution functions of known form whose parameters are in general unknown. The method detects and estimates the edge locations using a criterion derived by Bayesian theory. It chooses the number of edges and their locations according to the maximum *a posteriori* probability (MAP) principle. We provide results that demonstrate its performance on synthesized and real images.

### I. INTRODUCTION

A common task in image processing is edge detection. This problem has been widely researched, and there is a broad range of methods that can be used to resolve it satisfactorily [8]. Most of the work, however, has been devoted to the analysis of one-dimensional (1-D) images. In many practical applications including medical imaging, robotics, satellite technology, and industrial inspection, there is a need to process vector images, that is, multidimensional images whose pixels are represented by vectors. The edge detection in such images has been considerably less frequently addressed. Some references related to it include [3]–[6], [9], and [10].

We approach this problem by using Bayesian theory. As a principle for detecting edges, we exploit the MAP criterion, which selects the most probable hypothesis regarding the number of edges and their locations, given the observed image. The models associated with the hypotheses are described by parametric probability density functions. Their parameters are considered nuisance and are integrated out. The proposed criterion is a penalized likelihood function that is composed of two terms. The first term is the likelihood function that decreases monotonically as the number of hypothesized edges increases. The second term is the penalty function which, as opposed to the likelihood, increases monotonically as the number of edges increases, thus penalizing for using nonessential additional edges in modeling the image. The MAP solution is the one that minimizes the criterion.

An important feature of the proposed procedure is that no setting of thresholds is required for the edge detection. The procedure has been tested on synthesized and real magnetic resonance (MR) images. The results show that the number of edges as well as their locations are estimated with high accuracy even for low contrast-to-noise ratios (CNR's).

The paper is organized as follows. In the next section we formulate the problem, and in Section III we provide the criterion for edge detection. The details of its derivation are given in the Appendix. In Section IV, we present results, and in Section V some brief conclusions.

Manuscript received October 8, 1995; revised January 20, 1997. This work was supported by the NSF under Award 9506743. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. William E. Higgins.

The authors are with the Department of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350 USA (e-mail: djuric@sbee.sunysb.edu).

Publisher Item Identifier S 1057-7149(97)07895-0.



Fig. 1. A row of N pixels. The arrows point to the N+1 locations of possible boundaries.



Fig. 2. (a) Mask for detecting vertical boundaries. (b) Mask for detecting horizontal boundaries.

#### II. PROBLEM STATEMENT

A realization  $\{\mathbf{y}\}$  of a *q*-dimensional stochastic vector field  $\{\mathbf{Y}\}$  is observed over a set of pixel sites on an  $N_1 \times N_2$  image lattice.  $\{\mathbf{Y}\}$ is assumed to be a spatially quasistationary process in the sense that it is locally wide sense stationary. The random vector  $\mathbf{Y}_{ij}$  from the *i*th row and *j*th column, where  $(i, j) \in \mathbb{Z}_{N_1 \times N_2}$  with  $\mathbb{Z}_{N_1 \times N_2}$  being the set of pixel sites, is generated by the probability distribution function  $f(\mathbf{y}_{ij}|\boldsymbol{\theta}_{ij})$  whose parameters  $\boldsymbol{\theta}_{ij}$  are unknown. As one moves across the image, the parameters of the generating distribution function may change. It is assumed that these changes are abrupt as they reflect discontinuities of local image properties. The sites where the changes occur, which are referred to as edges, are unknown.

The generating probability distribution functions of the random vectors  $\mathbf{Y}_{ij}$ ,  $f(\mathbf{y}_{ij}|\boldsymbol{\theta}_{ij})$  are multivariate Gaussian whose parameters  $\boldsymbol{\theta}_{ij}$  are unknown. The parameters of each distribution are the mean vector  $\boldsymbol{\mu}_{ij}$  and the covariance matrix  $\boldsymbol{\Sigma}_{ij}$ . We restrict the covariance matrix to be diagonal, but not necessarily a multiple of the identity matrix. In other words

$$f(\mathbf{y}_{ij}|\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}) = \frac{1}{(2\pi)^{q/2} |\boldsymbol{\Sigma}_{ij}|^{1/2}} \\ \cdot \exp\left[-\frac{1}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_{ij})^T \boldsymbol{\Sigma}_{ij}^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_{ij})\right]$$
(1)

where  $\mathbf{y}_{ij}$  and  $\boldsymbol{\mu}_{ij}$  are q-dimensional vectors and  $\boldsymbol{\Sigma}_{ij}$  is a  $q \times q$  matrix, where  $\boldsymbol{\Sigma}_{ij} = \text{diag}(\sigma_{1_{ij}}^2, \sigma_{2_{ij}}^2, \cdots, \sigma_{q_{ij}}^2)$ .

Given the above assumptions, the problem is to detect all the edges and estimate their locations.

### III. CRITERION FOR DETECTION OF EDGES AND ESTIMATION OF THEIR LOCATION

We assume that there are two types of edges, horizontal and vertical. The horizontal edges are located at sites between pixels from the same column. If there is an edge at the site  $b^{(h)}$ , this implies that the generating distribution functions of the pixel data on each side of the edge are different. The vertical edges are defined similarly at sites between pixels from the same row and are denoted by  $b^{(v)}$ . As an example, in Fig. 1 we show one row with N pixels and the possible locations of the edges.

The edge locations in the *i*th row and *j*th column are denoted by the vectors  $\mathbf{b}_{\tilde{m}_i}^{(v)}$  and  $\mathbf{b}_{\tilde{m}_j}^{(h)}$ , respectively, where  $\tilde{m}_i$  and  $\tilde{m}_j$  are their individual lengths. More specifically, if the *i*th row of the image has  $\tilde{m}_i^{(v)}$  edges and their locations are denoted  $b_{k\tilde{m}_i}$  for

#### 1057-7149/97\$10.00 © 1997 IEEE

 $k = 1, 2, \cdots, \tilde{m}_i$ , then

$$\mathbf{b}_{\tilde{m}_{i}}^{(v)} = [b_{1\tilde{m}_{i}}^{(v)} \ b_{2\tilde{m}_{i}}^{(v)} \cdots \ b_{\tilde{m}_{i}\tilde{m}_{i}}^{(v)}]^{T}$$
(2)

where  $b_{k\bar{m}_i}^{(v)} < b_{l\bar{m}_i}^{(v)}$  for k < l and  $b_{k\bar{m}_i}^{(v)} \in \{k + 1, k + 2, \cdots, N_2 - \tilde{m}_i^{(v)} + k\}$ . The notation for horizontal edges is analogous. Note that we tacitly assume that the edge vectors may be of zeroth length, which is the case of no edges in the rows or columns.

To find the optimal estimates of the vector lengths and their elements, we use the Bayesian theory. If, as a criterion for best estimates, we adopt the maximum joint *a posteriori* probability, for the MAP estimates of the vertical edges we can write

$$[\hat{m}_{i}^{(v)}, \, \hat{\mathbf{b}}_{\tilde{m}_{i}}^{(v)}] = \arg \left\{ \max_{m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)}} \{ p(m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)} | \{\mathbf{y}\}) \} \right\}$$
(3)

where  $p(m_i^{(v)}, \mathbf{b}_{m_i}^{(v)} | \{\mathbf{y}\})$  is the joint *a posteriori* probability mass of  $m_i^{(v)}$  edges at the  $\mathbf{b}_{m_i}^{(v)}$  sites in the *i*th row. In (3),  $m_i^{(v)} \in \{0, 1, \dots, M_i^{(v)}\}$ , where  $M_i^{(v)}$  is the maximum number of edges in the *i*th row, and  $\mathbf{b}_{m_i}^{(v)} \in \mathcal{B}_{m_i}$ , where  $\mathcal{B}_{m_i}$  is the set of all possible vectors of length  $m_i$  whose structure is given by (2). An analogous expression holds for the horizontal edges. For convenience, we rewrite (3) in the following form:

$$[\hat{m}_{i}^{(v)}, \, \hat{\mathbf{b}}_{\hat{m}_{i}}^{(v)}] = \arg\left\{\min_{\substack{m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)}}} \{-\ln \, p(m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)}|\{\mathbf{y}\})\}\right\}.$$
 (4)

The derivation of the estimator based on (3) is rather lengthy, and all its details are omitted here. Its main steps are, however, provided in the Appendix. The obtained criterion for selection of the optimal number of edges and estimation of their locations can be expressed as

$$[\hat{m}_{i}^{(v)}, \, \hat{\mathbf{b}}_{\hat{m}_{i}}^{(v)}] = \arg\left\{\min_{m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)}} \left\{d(m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)}) + c(m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)})\right\}\right\}$$
(5)

where  $d[m_i^{(v)}, \mathbf{b}_{m_i}^{(v)}]$  and  $c[m_i^{(v)}, \mathbf{b}_{m_i}^{(v)}]$  are given by

$$d(m_i^{(v)}, \mathbf{b}_{m_i}^{(v)}) = \sum_{k=1}^{m_i^{(v)}+1} \frac{n_{i_k} - 2}{2} \sum_{l=1}^q \ln \hat{\sigma}_{l_{i_k}}^2$$
(6)

1...

and

$$c(m_i^{(v)}, \mathbf{b}_{m_i}^{(v)}) = [m_i^{(v)} + 1]q \ln \frac{K_1 K_2}{2\sqrt{2}\pi} + \sum_{k=1}^{m_i^{(0)}+1} q \ln n_{i_k} + \sum_{k=1}^{m_i^{(v)}} \ln \frac{N_2 - k}{k}.$$
(7)

In these expressions, some of the variables have the following meaning: q is the dimension of the vector field,  $n_{i_k}$  is the number of pixels in the kth segment of the *i*th row,  $\hat{\sigma}_{l_{i_k}}^2$  is the estimated noise variance from the kth segment in the *i*th row of the *l*th image,  $N_2$  is the number of pixels in a row,  $K_1$  and  $K_2$  are constants whose values depend on our prior knowledge about the vector field (see the Appendix). Note that  $d(m_i^{(v)}, \mathbf{b}_{m_i}^{(v)})$  is a term that depends on the data  $\{\mathbf{y}\}_i$  and that it decreases as the number of hypothesized edges increases. This is so because by allowing for more segments in the row, we can fit the observed data better. On the other hand, the term  $c(m_i^{(v)}, \mathbf{b}_{m_i}^{(v)})$  can be considered as a penalty because it increases as the number of hypothesized edges increases.

So, to find the vertical edges, we have to find the optimal solutions  $\hat{m}_i^{(v)}$ ,  $\hat{\mathbf{b}}_{m_i}^{(v)}$  for every  $i \in \{1, 2, \dots, N_2\}$  according to (5). For the horizontal edges, an equivalent expression is used. Clearly, the search for the best solution is computationally intensive since there is a large

number of possible combinations of edge locations. However, we can cope with this issue in several ways. First, the computation time can be dramatically decreased if the processing of the rows and the columns is parallelized. Second, there are two approaches that can decrease the computational intensity for processing each row/column. One is dynamic programming, which executes the optimization in (5) efficiently [1], [2]. The other is preprocessing of the vector image by two-dimensional (2-D) masks of the type shown in Fig. 2 [11]. For example, if the mask on the left in Fig. 2 is run for the *i*th row of the *q* images and if a potential edge is declared at the *j*th site when

$$\max\left[z_{ij}^{(1)}, \, z_{ij}^{(2)}, \, \cdots, \, z_{ij}^{(q)}\right] > \gamma \tag{8}$$

where  $z_{ij}^{(l)}$ ,  $l = 1, 2, \dots, q$  is the output of the mask when it is centered between the (j - 1)st and *j*th pixel of the *l*th image, and  $\gamma$  is an appropriately chosen threshold, then we obtain a set of edge candidates that can be tested via (5). Thus, instead of checking all the possible hypotheses, we are only examining a small subset of them.

The proposed procedure has a restriction that needs to be addressed. In its derivation, we have assumed that the complete vector image minus any row is independent of that row. Since this is an unrealistic assumption, we would like to relax it. One way to achieve this is to apply the results obtained by the masks in Fig. 2. Namely, first we run the masks through the rows and columns of all the images. By doing so, a map of potential edges is constructed. Second, we search for the MAP solution according to (5) by considering all the adjacent pixels to the processed row (or column) provided that between them and the row there are no horizontal (vertical) edges. Note that this inclusion does not have to be restricted to pixels from the adjacent rows (or columns) only. Also, in implementing (5) everything remains unchanged except that 1) the number of pixels  $n_{i_k}$  is different and should be carefully counted, and 2) in estimating  $\hat{\sigma}_{l_{i,i}}^2$  we exploit the adjacent pixels. This scheme yields an improved overall performance because by assembling more pixels for the row and column calculations, more contextual information is used and most of all, the approximation applied in the derivation becomes more accurate.

In summary, we suggest that the proposed procedure is implemented in two steps, as follows.

- Detection of potential edges by using the masks displayed in Fig. 2 and (8).
- 2) MAP estimation of the number of edges and their locations for each row and column by implementing (5). In processing the rows and columns, the edge candidates obtained in step one are tested. The tests include all the adjacent pixels that are not separated from the processed rows or columns by edges found in step one.

# IV. SIMULATION RESULTS

To test the performance of the proposed method, we conducted three experiments. In the first we generated 100 realizations of a subset of a three-dimensional (3-D) vector image, and in the second, we synthesized MR vector images. As performance measures for the algorithm, we used the following:

- 1) probability of false alarm (detecting a false edge);
- probability of missing an edge (not detecting an existing edge); and
- 3) mean squared errors (MSE's) of the edge location estimates.
- In all the experiments,  $K_1 = 256$ ,  $K_2 = 4.8$ , and  $\gamma = 14$ .

The generated subset in the first experiment represented three rows that do not have horizontal edges. The noiseless rows are shown in Fig. 3. As can be seen, there are m = 7 edges and their correct













Fig. 4. (a) First noisy row. (b) Second noisy row. (c) Third noisy row.



Fig. 5. First row: Three synthesized noisy MR images. The second and third rows represent results obtained without mask and with mask, respectively. From left to right: Estimated horizontal edge map, estimated vertical edge map, and estimated edge map.

locations are  $b_{17} = 20$ ,  $b_{27} = 40$ ,  $b_{37} = 50$ ,  $b_{47} = 70$ ,  $b_{57} = 80$ ,  $b_{67} = 90$ , and  $b_{77} = 100$ . The rows have 128 pixels. It should be observed that  $b_{17}$ ,  $b_{47}$ ,  $b_{57}$ , and  $b_{67}$  are present in each of the three images. On the other hand,  $b_{27}$ ,  $b_{37}$ , and  $b_{77}$  are only shown in two of them. Also,  $b_{57}$  and  $b_{67}$  are set close to each other so that we can test the performance of our algorithm when there are shorter segments. The noisy rows are displayed in Fig. 4. The CNR is equal to  $\frac{30}{15}$ . Note that we define the CNR as CNR = min { $|\mu_l - \mu_k|/\sigma$ }, where  $\mu_l$  and  $\mu_k$  are the intensity levels of the pixels on two different sides of an edge, and  $\sigma$  is the noise standard deviation. Clearly, the profiles from Fig. 4 show that it is not trivial to detect the number of edges and estimate their locations from the observed data.

We present the statistical results in Tables I and II. Excellent performance was obtained in detecting and estimating  $b_{17}$ ,  $b_{37}$ ,  $b_{47}$ , and  $b_{57}$ . The edges  $b_{27}$  and  $b_{67}$  were missed three times, and  $b_{77}$ , seven times in 100 trials. We had perfect estimation for  $b_{17}$ ,  $b_{37}$ , and  $b_{47}$ . The MSE was the largest for  $b_{77}$ . In Table II, we present the overall results of the detected and missed edges as well as the false alarms. It is interesting to note that it is much more likely to miss an edge than to produce a false alarm.

In the second experiment, the size of the three synthesized MR images (q = 3) was 256 × 256. The shapes of the various tissues

TABLE I Performance Results of the First Experiment when the CNR = 2/1

	<b>b</b> <sub>1,7</sub>	b <sub>2,7</sub>	b <sub>3,7</sub>	b <sub>4,7</sub>	b <sub>5,7</sub>	b <sub>6,7</sub>	b <sub>7,7</sub>
detected boundaries	100	97	100	100	100	97	93
missed boundaries	0	3	0	0	0	3	7
location MSE	0.0	0.14	0.0	0.0	0.05	0.02	0.23

TABLE II Total Number of Detected and Missed Boundaries as well as False Alarms for CNR = 2/1

detected boundaries	missing boundaries	false alarm
687	13	1

were obtained from a hand-segmented MR image. The parameters of the different tissues are shown in Table III.



Fig. 6. First row: Three real MR images. Second row from left to right: Estimated horizontal edge map, estimated vertical edge map, and estimated edge map (no mask used).

The three noisy images and the detection results are shown in Fig. 5. The first two images in the second row present the estimated vertical and horizontal edge maps, respectively, whereas the rightmost image displays all the detected edges. These edge maps were obtained without the use of mask. The bottom row provides the results of the procedure when it used the mask. Clearly, the algorithm was able to preserve the fine details of the MR images in either case.

Finally in the third experiment, we show results obtained from three real MR images showed in the top row of Fig. 6. In the bottom row we display the estimated horizontal, vertical, and overall edge maps, respectively. For visual comparison, in Fig. 7 we provide the results obtained by the vector gradient approach from [6] for the synthesized and real images in Figs. 5 and 6, respectively.

### V. CONCLUSIONS

We have presented an algorithm for edge detection of noisy vector fields. The algorithm is derived by using Bayesian theory, and it yields the joint MAP solution of the number of edges and their locations. Its final form has two terms, one of which is a data term and represents a measure of the modeling error, and the other is a penalty that quantifies the complexity of the model. The best solution is the one that minimizes the sum of the two terms. The algorithm can be implemented efficiently by processing the raws and the columns simultaneously. Also, to reduce the computational intensity, the method can be combined with simple edge detectors that provide a set of edge candidates that are subsequently tested. The performance of the algorithm was examined by computer simulations, which yielded very good results.

TABLE III PARAMETERS OF THE SYNTHESIZED IMAGES

	White Matter	Gray Matter	Ventricles	Bone	Else	
$\mu_1$	140	95	243	5	210	
$\sigma_1$	7	7	5	1.5	5	
$\mu_2$	130	100	90	5	200	
$\sigma_2$	5	6	5	1.5	5	
$\mu_3$	170	200	150	5	230	
$\sigma_3$	5	7	7	1.5	5	

## APPENDIX

Here we show the steps in deriving the main expressions (5)–(7). To simplify the presentation, consider the problem of finding the edges in the *i*th row only. The solution that maximizes  $p(m_i^{(v)}, \mathbf{b}_{m_i}^{(v)} | \{\mathbf{y}\})$  in (3) is identical to the one obtained from

$$[\hat{m}_{i}^{(v)}, \, \hat{\mathbf{b}}_{\hat{m}_{i}}^{(v)}] = \arg \left\{ \min_{\substack{m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)}}} \{ -\ln f(\{\mathbf{y}\} | m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)}) \\ -\ln p(m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)}) \} \right\}$$
(9)

where  $f({\mathbf{y}}|m_i^{(v)}, \mathbf{b}_{m_i}^{(v)})$  is the density function of the vector image data given the edges in the *i*th row, and  $p(m_i^{(v)}, \mathbf{b}_{m_i}^{(v)})$  is the *a priori* 



Fig. 7. Edge detection results obtained for the images in Figs. 5 and 6, respectively, using a vector gradient approach.

probability mass function of the number of edges and their locations. The *a priori* probability mass function  $p(m_i^{(v)}, \mathbf{b}_{m_i}^{(v)})$  can be written as

$$p(m_i^{(v)}, \mathbf{b}_{m_i}^{(v)}) = p(\mathbf{b}_{m_i}^{(v)} | m_i^{(v)}) p(m_i^{(v)})$$
(10)

where  $p(m_i^{(v)})$  is the *a priori* probability of  $m_i^{(v)}$  edges. For the prior  $p(m_i^{(v)})$  we adopt the uniform mass function. If the maximum number of possible edges in the row is  $M_i^{(v)}$ , then

$$p(m_i^{(v)}) = \frac{1}{M_i^{(v)}}.$$
(11)

For the conditional prior  $p(\mathbf{b}_{m_i}^{(v)}|m_i^{(v)})$  we also take up the uniform probability mass function. Since the number of combinations of  $m_i^{(v)}$  edges in  $N_2 - 1$  locations can easily be found, it follows that

$$p(\mathbf{b}_{m_i}^{(v)}|m_i^{(v)}) = \frac{m_i^{(v)}!(N_2 - m_i^{(v)} - 1)!}{(N_2 - 1)!}.$$
 (12)

So, from (10)-(12), we conclude that (9) can be approximated by

$$[\hat{m}_{i}^{(v)}, \hat{\mathbf{b}}_{\hat{m}_{i}}^{(v)}] = \arg \left\{ \min_{m_{i}^{(v)}, \mathbf{b}_{m_{i}}^{(v)}} \left\{ -\ln f(\{\mathbf{y}\} | m_{i}^{(v)}, \mathbf{b}_{m_{i}}^{(v)}) - \sum_{k=1}^{m_{i}^{(v)}} \ln k + \sum_{k=1}^{m_{i}^{(v)}} \ln (N_{2} - k) \right\} \right\}.$$
(13)

Now we examine the term  $\ln f({\mathbf{y}}|m_i^{(v)}, \mathbf{b}_{m_i}^{(v)})$ . To simplify the presentation, first we assume that

$$\ln f(\{\mathbf{y}\}|m_i^{(v)}, \mathbf{b}_{m_i}^{(v)}) = \ln f(\{\mathbf{y}\}_i|m_i^{(v)}, \mathbf{b}_{m_i}^{(v)}) + \ln f(\{\mathbf{y}\}_{(-i)}|m_i^{(v)}, \mathbf{b}_{m_i}^{(v)})$$
(14)

where  $\{\mathbf{y}\}_i$  denotes the vectors from the *i*th row, and  $\{\mathbf{y}\}_{(-i)}$  represents the complete vector image without the *i*th row. If, in addition

$$\ln f(\{\mathbf{y}\}_{(-i)}|m_i^{(v)}, \mathbf{b}_{m_i}^{(v)}) = \ln f(\{\mathbf{y}\}_{(-i)})$$
(15)

this entails that (13) can be expressed as

$$(\hat{m}_{i}^{(v)}, \, \hat{\mathbf{b}}_{\tilde{m}_{i}}^{(v)}) = \arg \left\{ \min_{\substack{m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)}}} \left\{ -\ln f[\{\mathbf{y}\}_{i} | m_{i}^{(v)}, \, \mathbf{b}_{m_{i}}^{(v)}] - \sum_{k=1}^{m_{i}^{(v)}} \ln k + \sum_{k=1}^{m_{i}^{(v)}} \ln (N_{2} - k) \right\} \right\}.$$

$$(16)$$

The assumptions (14) and (15) are rather strong, since they imply independence between the observed vectors from the *i*th row and the

rest of the vector image, given the number of vertical edges and their locations in the *i*th row. However, as discussed in this work, there is a way to relax them.

Next, from our assumptions we may write

$$\ln f(\{\mathbf{y}\}_i | m_i^{(v)}, \mathbf{b}_{m_i}^{(v)}) = \sum_{k=1}^{m_i^{(v)}+1} \ln f(\{\mathbf{y}\}_{i_k})$$
(17)

where  $f(\{\mathbf{y}\}_{i_k})$  is the marginal probability distribution function of the vectors  $\mathbf{y}_{ij}$  between  $b_{(k-1)m_i}^{(v)}$  and  $b_{km_i}^{(v)}$  (by assumption  $b_{0m_i} = 1$ , and  $b_{(m_i+1)m_i} = N_2 + 1$ ). So, we need to determine  $\ln f(\{\mathbf{y}\}_{i_k})$ . We use

$$f(\{\mathbf{y}\}_{i_k}) = \int_{\boldsymbol{\Theta}_{i_k}} f(\{\mathbf{y}\}_{i_k} | \boldsymbol{\theta}_{i_k}) f(\boldsymbol{\theta}_{i_k}) \, d\boldsymbol{\theta}_{i_k}$$
(18)

where

$$f(\{\mathbf{y}\}_{i_k}|\boldsymbol{\theta}_{i_k}) = \prod_{\substack{j=b_{(k-1)m_i}^{(v)}}}^{b_{km_i}^{(v)}-1} f(\mathbf{y}_{ij}|\boldsymbol{\theta}_{i_k}).$$
(19)

Here,  $\boldsymbol{\theta}_{i_k}$  is the set of parameters of the multivariate Gaussian function given by (1),  $\boldsymbol{\Theta}_{i_k}$  is the parameter space of  $\boldsymbol{\theta}_{i_k}$ , and  $f(\boldsymbol{\theta}_{i_k})$ , the prior probability distribution function of  $\boldsymbol{\theta}_{i_k}$ . We assume that the prior mean and covariance are independent, or

$$f(\boldsymbol{\theta}_{i_k}) = f(\boldsymbol{\mu}_{i_k}) f(\boldsymbol{\Sigma}_{i_k}).$$
(20)

If the range of intensity values of each element of  $\mu_{i_k}$  is  $[x_{\min}, x_{\max}]$ , then for  $f(\mu_{i_k})$  we adopt

$$f(\boldsymbol{\mu}_{i_k}) = \frac{1}{K_1^q} \tag{21}$$

where  $K_1 = x_{\max} - x_{\min}$ . For  $f(\Sigma_{i_k})$  we write

$$f(\boldsymbol{\Sigma}_{i_k}) = \frac{1}{K_2^q} \prod_{l=1}^q \frac{1}{\sigma_{l_{i_k}}}, \qquad \sigma_{l_{i_k}} \in [\sigma_{\min}, \sigma_{\max}].$$
(22)

where  $K_2 = \ln \sigma_{\max} - \ln \sigma_{\min}$ .

To evaluate the integral in (18), we use (19)–(22). Under the given assumptions, the final result cannot be given in a closed form. Therefore, we resort to an approximation that allows for an easier computation of (18), and a result that has an interesting interpretation. If we expand the logarithm of  $f(\{\mathbf{y}\}_{i_k} | \boldsymbol{\theta}_{i_k}) f(\boldsymbol{\theta}_{i_k})$  around the maximum likelihood estimate of  $\boldsymbol{\theta}_{i_k}$ ,  $\hat{\boldsymbol{\theta}}_{i_k}$ , we obtain [7]

$$\ln \left[ f(\{\mathbf{y}\}_{i_k} | \boldsymbol{\theta}_{i_k}) f(\boldsymbol{\theta}_{i_k}) \right] \\ \simeq \ln f(\{\mathbf{y}\}_{i_k} | \hat{\boldsymbol{\theta}}_{i_k}) + \ln f(\hat{\boldsymbol{\theta}}_{i_k}) \\ - \frac{1}{2} \left( \boldsymbol{\theta}_{i_k} - \hat{\boldsymbol{\theta}}_{i_k} \right)^T \mathcal{H}_{i_k} (\boldsymbol{\theta}_{i_k} - \hat{\boldsymbol{\theta}}_{i_k})$$
(23)

where  $\mathcal{H}_{i_k}$  is the Hessian matrix [7] whose size is  $2q \times 2q$ , and which can be approximated by

$$\mathcal{H}_{i_{k}} = \operatorname{diag}\left(\frac{n_{i_{k}}}{\hat{\sigma}_{1_{i_{k}}}^{2}}, \frac{n_{i_{k}}}{2\hat{\sigma}_{1_{i_{k}}}^{4}}, \frac{n_{i_{k}}}{\hat{\sigma}_{2_{i_{k}}}^{2}}, \frac{n_{i_{k}}}{2\hat{\sigma}_{2_{i_{k}}}^{4}}, \cdots, \frac{n_{i_{k}}}{\hat{\sigma}_{q_{i_{k}}}^{2}}, \frac{n_{i_{k}}}{2\hat{\sigma}_{q_{i_{k}}}^{4}}\right).$$
(24)

When the approximation of  $f(\{\mathbf{y}\}_{i_k}|\boldsymbol{\theta}_{i_k})f(\boldsymbol{\theta}_{i_k})$  by (23) is substituted in (18) and the integration carried out, we obtain

$$\ln f(\{\mathbf{y}\}_{i_k}) \simeq \ln f(\{\mathbf{y}\}_{i_k} | \hat{\boldsymbol{\theta}}_{i_k}) + \ln f(\hat{\boldsymbol{\theta}}_{i_k}) + q \ln (2\pi) - \frac{1}{2} \sum_{l=1}^{q} \ln \frac{n_{i_k}^2}{2\hat{\sigma}_{l_{i_k}}^6}.$$
 (25)

Since

$$\ln f(\{\mathbf{y}\}_{i_k}|\hat{\boldsymbol{\theta}}_{i_k}) = -\frac{n_{i_k}}{2} \sum_{l=1}^q \ln \left(2\pi\hat{\sigma}_{l_{i_k}}^2\right) - \frac{qn_{i_k}}{2}$$
(26)

we rewrite (25) as

$$\ln f(\{\mathbf{y}\}_{i_k}) \simeq -\frac{n_{i_k} - 2}{2} \sum_{l=1}^{q} \ln \hat{\sigma}_{l_{i_k}}^2 - \frac{q(n_{i_k} - 2)}{2} \ln (2\pi) - \frac{qn_{i_k}}{2} - q \ln(C_1 C_2) - q \ln n_{i_k} + \frac{q}{2} \ln 2.$$
(27)

From (17) and (27) and after dropping irrelevant terms, we can finally write the criterion (16) as shown in (5) and (6).

#### References

- [1] R. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.
- [2] N. Burobin, V. Mott'l, and I. Muchnik, "An algorithm for detection of the instant of multiple change of properties of a random process based on the dynamic programming method," in *Detection of Changes in Random*

Processes, L. Telksnys, Ed. New York: Optimization Software, Inc., 1986.

- [3] A. Cumani, "Edge detection in multispectral images," CVIGP: Graph. Models Image Process., vol. 53, pp. 40–51, 1991.
- [4] C. J. Delcroix and M. A. Abidi, "Fusion of edge maps in color images," in Proc. SPIE-Int. Soc. Opt. Eng., 1988, vol. 1001, pp. 454–554.
- [5] J.-K. Fwu and P. M. Djurić, "Unsupervised vector image segmentation by a tree structure—ICM algorithm," *IEEE Trans. Med. Imag.*, vol. 15, pp. 871–880, 1996.
- [6] H.-C. Lee and D. R. Cok, "Detecting boundaries in a vector field," *IEEE Trans. Signal Processing*, vol. 39, pp. 1181–1194, 1991.
- [7] A. O'Hagan, Kendall's Advanced Theory of Statistics: Bayesian Inference, vol. 2B. New York: Wiley, 1994.
- [8] A. Rosenfeld and A. Kak, *Digital Picture Processing*. New York: Academic, 1981.
- [9] P. E. Trahanias and A. N. Venetsanopoulos, "Color edge detection using vector order statistics," *IEEE Trans. Image Processing*, vol. 2, pp. 259–264, 1993.
- [10] J. K. Udupa, "Multidimensional digital boundaries," CVGIP: Graph. Models Image Process., vol. 56, pp. 311–323, 1994.
- [11] Z. Wu and R. Leahy, "Image segmentation via edge contour finding: A graph theoretic approach," in *Proc. IEEE Comput. Soc. Conf. Computer Vision and Pattern Recognition*, 1992, pp. 613–619.