OPTIMAL SEGMENTATION OF SIGNALS AND ITS APPLICATION TO IMAGE DENOISING AND BOUNDARY FEATURE EXTRACTION

Tony X. Han[†]   *Steven Kay*[‡]   *Thomas S. Huang*[†]

† Beckman Institute and ECE Department
University of Illinois at Urbana-Champaign
405 N. Mathews Ave., Urbana, IL 61801, USA
{xuhan, huang}@ifp.uiuc.edu

‡ ECE Department
University of Rhode Island
4 East Alumni Ave, Kingston, RI 02881, USA
kay@ele.uri.edu

ABSTRACT

An optimal procedure for segmenting one-dimensional signals whose parameters are unknown and change at unknown times is presented. The method is maximum likelihood segmentation, which is computed using dynamic programming. In this procedure, the number of segments of the signal need not be known a priori but is automatically chosen by the Minimum Description Length rule. The signal is modeled as unknown DC levels and unknown jump instants with an example chosen to illustrate the procedure. This procedure is applied to image denoising and boundary feature extraction. Because the proposed method uses the global information of the whole image, the results are more robust and reasonable than those obtained through classical procedures which only consider local information. The possible directions for improvement are discussed in the conclusion.

1. INTRODUCTION

Signal segmentation is a very important problem that appears in many areas such as speech recognition, image feature extraction, edge detection, signal detection and communications [1]. The problem is stated as following. A data set $\{x[0], x[1], \ldots, x[N-1]\}$ is observed which is composed of N_s segments of differing statistics. With N_s segments there are $N_s - 1$ transition times composing the set $\{n_1, n_2, \ldots, n_{N_s-1}\}$. Additionally, each segment may depend on a set of unknown parameters $\boldsymbol{\theta}$. The most general segmentation problem is to estimate the number of segments N_s, the transition times n_i, and the unknown vector of parameters $\boldsymbol{\theta}_i$ for each segment.

There have been numerous attempts to solve the segmentation problem [2-9]. A statistically optimal approach (for large data records) is to find the maximum likelihood estimator (MLE) of the unknown parameters [10]. Unfortunately, due to the computational burden of implementation which grows exponentially with the number of segments, this has not been pursued. Suboptimal methods based on sequential estimation have been proposed to reduce the computation. The disadvantage is the reduced performance as well as the need for the setting of various thresholds. The choice of the thresholds tends to be ad hoc and so reliable segmentation cannot always be obtained.

In this paper we propose an MLE segmenter whose computational complexity is reduced drastically over a conventional implementation. This reduction is due to the use of dynamic programming (DP) [11]. The approach which is proposed has been motivated by the work of Bellman in fitting a piecewise linear model to a given curve [12]. Our work was originally implemented for segmentation of autoregressive process in [16]. It should also be mentioned that similar methods have been independently reported in [13] and [17].

The paper is organized as follows. Section 2 derives the DP formalism for solving the segmentation problem. Section 3 shows the application of DP to segment the signal modeled as unknown DC levels with unknown jump instants. In section 4 those results are applied to the problem of image denoising and in Section 5 they are applied to boundary feature extraction. Section 6 concludes the paper.

2. MAXIMUM LIKELIHOOD SEGMENTATION FOR ONE-DIMENSIONAL SIGNAL

2.1. Known Number of Segments

We wish to segment a discrete time series (one-dimensional signal) $\{x[0], x[1], \ldots, x[N-1]\}$ into N_s segments by choosing the set of transition times $\{n_1, n_2, \ldots, n_{N_s-1}\}$. To do so we assume that the ith segment is characterized by the PDF $p_i\left(x[n_{i-1}], \ldots, x[n_i - 1]; \boldsymbol{\theta}_i\right)$, where $\boldsymbol{\theta}_i$ is a vector of unknown parameters. Furthermore, each segment is assumed to be statistically independent of all other segments. With these assumptions the PDF of the data set can be written as

$$\prod_{i=1}^{N_S} p_i\left(x[n_{i-1}],\ldots,x[n_i-1];\boldsymbol{\theta}_i\right)\cdot \qquad (1)$$

Since by definition $n_0 \equiv 0$, $n_{N_S} \equiv N$, the MLE segmenter chooses $\left\{n_1,n_2,\ldots,n_{N_S-1},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_{N_S}\right\}$ which maximizes (1), or by defining

$$\mathbf{X}[i,j] = \left[\, x[i] \quad x[i+1] \quad \ldots \quad x[j]\,\right]^T$$

the MLE segmenter must maximize

$$\sum_{i=1}^{N_S} \ln p_i\left(\mathbf{X}[n_{i-1},n_i-1];\boldsymbol{\theta}_i\right)\cdot \qquad (2)$$

Note that to maximize (2) we need to
1)Choose a set of transition times; 2)Estimate $\boldsymbol{\theta}_i$ for each segment to yield $p_i\left(\mathbf{X}[n_{i-1},n_i-1];\hat{\boldsymbol{\theta}}_i\right)$; 3)Sum all log PDFs for all $p_i\left(\mathbf{X}[n_{i-1},n_i-1];\hat{\boldsymbol{\theta}}_i\right)$; 4)Repeat the calculation for every possible set of transition times to determine which one yields the maximum.

The difficulty is of course due to the exponential rate of increase of number of possible partitions with number of segments. To circumvent this we let

$$\Delta_i\left[n_{i-1},n_i-1\right] = -\ln p_i\left(\mathbf{X}[n_{i-1},n_i-1];\hat{\boldsymbol{\theta}}_i\right)$$

so that we wish to minimize

$$\sum_{i=1}^{N_S} \Delta_i\left[n_{i-1},n_i-1\right] \qquad (3)$$

over $\left\{n_1,n_2,\ldots,n_{N_S-1}\right\}$. (It is assumed that the MLE of $\boldsymbol{\theta}_i$ for the ith segment is easily found once the segment boundaries have been specified.) But (3) has the Markovian form and is readily solvable via DP. To do so let

$$I_k(L) = \min_{\substack{n_1,n_2\ldots n_{k-1}\\ n_0=0,n_k=L+1}} \sum_{i=1}^{k} \Delta_i\left[n_{i-1},n_i-1\right] \qquad (4)$$

which represents the minimum negative log likelihood function for k segment partitions of the data record $\left\{x[0],x[1],\ldots,x[L]\right\}$. Note there are some implicit restrictions on the transition times, which result from $1 \le n_1 < n_2 < \cdots < n_{k-1} \le L$. Now,

$$\begin{aligned}
I_k(L) &= \min_{\substack{n_{k-1}\\ n_k=L+1}} \min_{\substack{n_1,n_2\ldots n_{k-2}\\ n_0=0}} \sum_{i=1}^{k} \Delta_i\left[n_{i-1},n_i-1\right]\\
&= \min_{\substack{n_{k-1}\\ n_k=L+1}} \min_{\substack{n_1,n_2\ldots n_{k-2}\\ n_0=0}} \sum_{i=1}^{k-1} \Delta_i\left[n_{i-1},n_i-1\right] + \Delta_k\left[n_{k-1},n_k-1\right]\\
&= \min_{\substack{n_{k-1}\\ n_k=L+1}} \left[I_{k-1}\left(n_{k-1}-1\right) + \Delta_k\left[n_{k-1},n_k-1\right]\right]\\
&= \min_{n_{k-1}} \left[I_{k-1}\left(n_{k-1}-1\right) + \Delta_k\left[n_{k-1},L\right]\right]
\end{aligned}$$

where we have used (4). Including the constraints on the transition times

$$I_k(L) = \min_{k-1\le n_{k-1}\le L}\left[I_{k-1}\left(n_{k-1}-1\right) + \Delta_k\left[n_{k-1},L\right]\right] \quad (5)$$

for $L = k-1,k,\ldots,N-1$. The solution to our original problem occurs for $k = N_S$ and $L = N-1$. To begin the recursion we need to compute

$$I_1(L) = \Delta_1\left[n_0,L\right] = \Delta_1\left[0,L\right] = -\ln p_1\left(\mathbf{X}[0,L];\hat{\boldsymbol{\theta}}_1\right) \quad (6)$$

for $L = 0,1,\ldots,N-1$. The actual procedure embodied in (5) and (6) is as follows:
1) Compute (6) for $L = 0,1,\ldots,N-1$ and store the results in $I_1(L)$. This is the maximum likelihood for all data records from $n=0$ to $n=L$ or for all the one-segment "partitions".
2) According to (5) the optimal two-segment partitions as a function of data record length are found as

$$I_2(L) = \min_{1\le n_1\le L}\left[I_1\left(n_1-1\right) + \Delta_2\left[n_1,L\right]\right]\cdot$$

To compute this we need to compute $\Delta_2\left[n_1,L\right]$ for all $1 \le n_1 \le L$ (the lower limit is 1 to allow for a minimum one sample first segment). To this we add $I_1\left(n_1-1\right)$ which has already been found and stored in step 1.
3) We continue the procedure requiring at each step the computation of $\Delta_k\left[n_{k-1},L\right]$. Hence the computation goes up linearly with number of segments.

It should also be noted that due to the recursive nature of the DP solution the best $2,3,\ldots,N_S-1$ segmentations are found as a byproduct of the approach. This allows us to determine the best segmentation when the number of segments is not known a priori as described next.

2.2. Unknown Number of Segments
In order to determine the number of segments we employ the Minimum Description Length (MDL) [14]. If we were naively to choose the partition as the one which maximizes the likelihood, then we would always choose the maximum number of segments. This is because more parameters (the $\boldsymbol{\theta}_i$) are estimated as more segments are assumed, causing the likelihood function to monotonically increase with number of segments. The MDL as applied to our problem chooses the number of segments $\hat{N}_S$ as the value of k which minimizes for $k = 1,2,\ldots,S$

$$\text{MDL}(k) = -\ln \prod_{i=1}^{k} p_i\left(x[\hat{n}_{i-1}],\ldots,x[\hat{n}_i-1];\hat{\boldsymbol{\theta}}_i\right) + \frac{r_k}{2}\ln N\cdot(7)$$

Here S is the maximum number of segments, which is chosen by the user. $\{\hat{n}_1,\hat{n}_2,\ldots,\hat{n}_{k-1},\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\theta}}_2,\ldots,\hat{\boldsymbol{\theta}}_k\}$ is the MLE for the k-segment partition of the entire data record.

By definition $\hat{n}_0 = 0$, $\hat{n}_{N_s} = N$. Finally, r_k is the number of parameters estimated for an assumed k segments. If the dimension of $\hat{\boldsymbol{\theta}}_i$ is q_i, then

$$r_k = \sum_{i=1}^{k} q_i + k - 1 \qquad (8)$$

with the first term of (8) representing the estimated $\boldsymbol{\theta}_i$ and the second due to the $k-1$ transition times.

The first term of the MDL of (7) may be computed using DP as described in the previous section. This is because the maximum likelihood solution using DP also provides all the lower order solutions as well. Almost no extra computation is required when the number of segments is unknown.

3. ONE-DIMENSIONAL SIGNAL MODELING AND ITS SEGMENTATION

We model the one-dimensional signal as a multilevel DC signal, which jumps at the unknown transition times and is contaminated by white Gaussian noise (WGN). The signal is

$$s[n] = \sum_{i=1}^{N_s} A_i \left[u[n_{i-1}] - u[n_i] \right] \qquad (9)$$

where $n_0 = 0 < n_1 < n_2 < \cdots < n_{N_s-1} < n_{N_s} = N$ and $u[n]$ is the unit step sequence. Here N is the length of the signal. The contaminated signal we observed is

$$x[n] = s[n] + w[n] \qquad (10)$$

where $w[n]$ is WGN. Let $\mathbf{X} = \begin{bmatrix} x[0] & x[1] & \cdots & x[N-1] \end{bmatrix}^T$. The PDF of $\mathbf{X}$ is

$$p(\mathbf{X}; \mathbf{A}, \mathbf{n}) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{N_s} \left[\sum_{n=n_{i-1}}^{n_i-1} \left(x[n] - A_i \right)^2 \right] \right] (11)$$

where $\mathbf{A} = \begin{bmatrix} A_1 & A_2 & \cdots & A_{N_s} \end{bmatrix}^T$, $\mathbf{n} = \begin{bmatrix} n_1 & n_2 & \cdots & n_{N_s-1} \end{bmatrix}^T$ and σ^2 is the variance of the WGN. From equation (11), the MLE of A_i is

$$\hat{A}_i = \frac{1}{n_i - n_{i-1}} \sum_{n=n_{i-1}}^{n_i-1} x[n] \cdot \qquad (12)$$

Then

$$\Delta_i \left[n_{i-1}, n_i - 1 \right] = -\ln p_i \left(\mathbf{X}[n_{i-1}, n_i - 1]; \hat{\boldsymbol{\theta}}_i \right)$$
$$= \frac{1}{2\sigma^2} \sum_{n=n_{i-1}}^{n_i-1} \left(x[n] - \hat{A}_i \right)^2 + \frac{n_i - n_{i-1}}{2} \ln\left(2\pi\sigma^2\right) \qquad (13)$$

Since $\sum_{i=1}^{N_s} \left(n_i - n_{i-1} \right) = N$, we need only minimize

$$\sum_{i=1}^{N_s} \Delta_i \left[n_{i-1}, n_i - 1 \right] \qquad (14)$$

where $\Delta_i \left[n_{i-1}, n_i - 1 \right] = \sum_{n=n_{i-1}}^{n_i-1} \left(x[n] - \hat{A}_i \right)^2$. Comparing Equation (14) with Equation (4) and (5), we know that we

can use the method of DP to find $\hat{\mathbf{A}}$ and $\hat{\mathbf{n}}$ which maximize $p(\mathbf{X}; \mathbf{A}, \mathbf{n})$. Here, because the number of segments is unknown, we need to compute the $\text{MDL}(k)$ using equation (7). Then the estimation of the number of segments, $\hat{N}_s$ is the "valley" of value of k which globally minimizes the $\text{MDL}(k)$ over all possible segment numbers.

A computer simulation example is show in Figure 1. In this example, the multi-levels of the original signal is $\mathbf{A} = [1 \ 4 \ 2 \ 6 \ -2 \ 1 \ 5 \ 2]^T$. The corresponding transition times are $\mathbf{n} = [20 \ 50 \ 65 \ 100 \ 105 \ 110 \ 115]^T$. The DP estimate of the transition times is $\hat{\mathbf{n}} = [20 \ 50 \ 65 \ 100 \ 105 \ 110 \ 115]^T$, and the MLE of the multi-levels is $\hat{\mathbf{A}} = [1.14 \ 4.13 \ 1.45 \ 6.04 \ -2.49 \ 0.86 \ 4.94 \ 2.35]^T$. From this experiment we find that DP estimations of the transition times and number of segments are very accurate, even under the condition of low SNR. It is important to notice another advantage of the algorithm. At the end of the multiple DC level signal, the length of each DC level is just 5 samples. However, the DP algorithm can still detect the transition times very precisely. This motivates one to apply this algorithm to image denoising and edge detection.
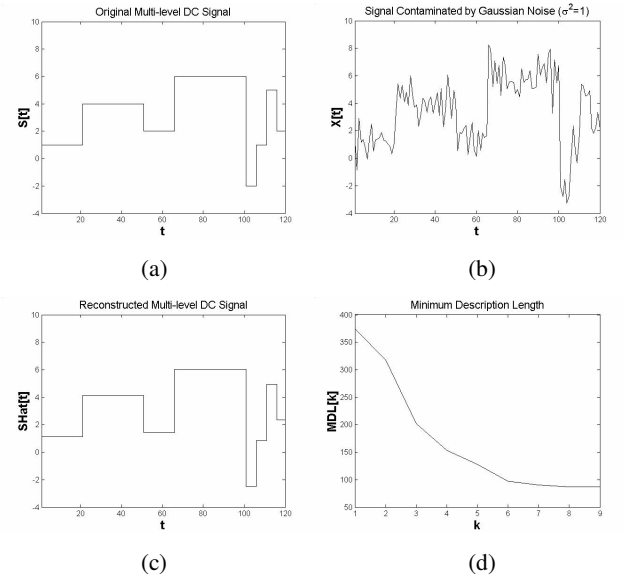

(a)

(b)

(c)

(d)

Figure 1. Segmentation and reconstruction of one-dimensional multi-level signal. (Unknown number of segments. (a)Original multi-level DC signal; (b)Observed signal contaminated by white Gaussian noise ($\sigma^2 = 1$); (c)Reconstructed multi-level DC signal; (d)Minimum Description Length of the segmentation.

4. APPLICATION TO IMAGE DENOISING

By generalizing the results of the DP segmentation and reconstruction of 1-D signal to 2-D signal, i.e the image, image denoising is done by 1) reconstruct 1d signal horizontally (note the result image $I_H(i,j)$) and vertically (note the result image $I_V(i,j)$); 2) Take the average of the two image by

$$I_D(i,j) = \sqrt{\frac{I_H(i,j)^2 + I_V(i,j)^2}{2}} . \quad (15)$$

The average SNR(db) of denoising results over many standard images compared with traditional methods are show in table 1. (Images are omitted due to space limit.)

σ^2 of WGN	100	400	900
SNR(db) of image with noise	22.38	16.34	12.80
SNR(db) after low pass filtering	17.80	17.34	16.62
SNR(db) after median filtering	20.43	18.78	16.97
SNR(db) of DP segmentation/reconstruction	23.90	19.69	17.29

Table 1. The SNR of the denoised images.

5. APPLICATION TO EDGE DETECTION

By detect the jump position of the 2-D image horizontally and vertically, this algorithm can find the global boundary feature, while the traditional algorithm extracting boundary feature by edge detection only consider the local response of a compact support filter. The boundary feature extraction results compared with Canny edge detector are shown in figure 2.

6. CONCLUSION

Maximum likelihood segmentation appears to work quite well for one-dimensional signals, but the computation is extremely intensive. When DP is used, the computational complexity is substantially reduced from exponentially to only linearly increasing with an increased number of segments. MDL is introduced to decide the optimal segment number in the sense of minimum description length. The AR/MA process, or other statistical models may be used to represent the image [15]. MLE segmentation based on these models can still be realized through DP [16].

7. REFERENCES

[1] S. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, Prentice-Hall, 1998.

[2] P. Prandoni, M. Goodwin, and M. Vetterli, "Optimal Segmentation for Signal Modelling and Compression," *ICASSP 97*, pp.465-471, 1997.

[3] S. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multi-band image segmentation," *IEEE Trans. PAMI*, vol.18 pp.884-900, 1996.

[4] E. Kolaczyk and R. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," *Annals of Statistics*, (in press), 2004.

[5] Thomas C. M. Lee, "Segmenting Images Corrupted by Correlated Noise," *IEEE Trans. PAMI* vol.20, pp.481-492 , 1998.

[6] Y. G. Leclerc, "Constructing simple stable descriptions for image partitioning," IJCV, vol 3(1), pp.73-102, 1989.

[7] M. Basseville and A. Benveniste, "Sequential Detection of Abrupt Changes in Spectral Characteristics of Digital Signals," *IEEE Trans. in Info. Theory*, Sept, pp.709-724, 1983.

[8] R. Andre-Obrecht, "A New Statistical Approach for the Automatic Segmentation of Continuous Speech Signals*," IEEE Trans. on ASSP, vol.36*, pp.29-40, 1988.

[9] A.S. Willsky, "A Survey of Design Methods for Failure Detection in Dynamic Systems," *Automatica*, vol.12, pp.601-611, 1976.

[10] S. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*, Prentice-Hall, 1993.

[11] R. E. Larson and J. L. Castie, *Principles of Dynamic Programming, vol. I, II*, Marcel Dekker Inc., NY, 1982.

[12] R. Bellman and S. Dreyfus, *Applied Dynamic Programming*, Princeton Univ. Press, Princeton, NJ, 1962.

[13] N. Burobin, V. Mottl, and I. Muchnik, "An Algorithm of Detection of Multiple Change of Properties of Random Process Based on the Dynamic Programming Method," in Detection of Change in Random Processes, Optimization Software Inc., NY, 1986.

[14] J. Rissanen, "Modeling by Shortest Data Description," *Automatica*, vol. 14, 1978, pp.465-471.

[15] S. Kay, Modern Spectral Estimation, Prentice-Hall, Englewood Cliffs, NJ, 1988

[16] S. Kay, "Optimal Segmentation of Time Series Based on Dynamic Programming," Unpublished notes, 1989.

[17] T. Svendsen and F. K. Soong, "On the Automatic Segmentation of Speech Signals," *ICASSP*, pp. 77-80, 1987.
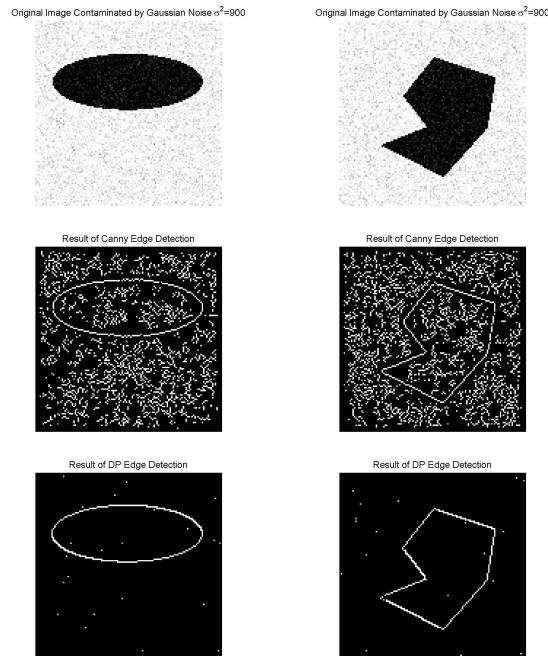
Figure 2 The boundary feature extraction results. First row: The original images contaminated by WGN; Second row: the results of Canny edge detection; The third row: the results of optimal DP segmentation combined with MDL detection.