ON SEQUENTIAL ON-LINE OUTLIER DETECTION AND A LINESCAN APPLICATION

J. Vihonen, T. Ala-Kleemola, R. Kerminen, J. Jylhä, A. Visa

Tampere University of Technology, Institute of Signal Processing
Korkeakoulunkatu 1, P.O. Box 553, FI-33101 Tampere, Finland

ABSTRACT

Industrial quality monitoring is increasing rapidly, and challenging signal environments with requirement of steady performance pose conflicting demands to on-line tests. The sequential probability ratio test (SPRT) and the Kalman filter (KF) are proposed as two tools for detection and recognition-oriented signal processing. A modified sequential test is suggested and applied to a linescan problem.

1. INTRODUCTION

Systems ranging from mechanical vibration monitoring to elegant sensor life-cycle management and fault detection have become increasingly popular in industry. Many of these applications adopt and benefit from on-line detection of a change. Linescan camera applications have also established a firm foothold in several civilian and military applications.

Excellent techniques exist for the detection of changes under the assumption of statistical independence. Assuming that the measured signal does not admit the independence assumption, the appearance of increased-variance observations should be distinguished as transient state. For example in papermaking, variations in mass distribution of pulp and other raw materials result in thick or thin spots or even holes, giving dependencies some structural form. As it might be expected, accurate segmentation and decomposition of such local signal characteristics is of interest. Two tools, optimum in their own categories, are proposed for this task. A wide-ranging image change detection survey can be found in [1].

The Page's detector [2] is based on the idea of Wald's SPRT [3]. These methods do not need to determine the number of measurements for the test in advance. In addition to simplicity, a notable property of SPRT is that with given fixed values of error probabilities it is optimal for testing between two independent distributions in the sense that it minimizes average run length [4]. Appealingly, Page's test and SPRT work well in practice for dependent sequences.

Stochastic local process transitions can be captured through KF residual analysis. For hypothesis testing, this is an attractive property since complete process-characterizing transient state models are not necessarily required. Assuming that the true process state model is known, the filter is optimal in the sense that it minimizes the Bayesian mean squared error for each new estimation [5]. The filter is also the optimal linear minimum mean squared error estimator, if the Gaussian statistics are invalid. A recent KF residual treatment can be found in [6].

With simple deduction, it becomes obvious that stepwise process tracking capitalizes unexpected, ”hard” transitions. Slow or ”soft” changes remain easily undetected. Therefore, both the KF and SPRT have differing advantages. Adopting the expertise of these methods to a single detector is proposed. Here, the residual check is embedded in SPRT, which is otherwise treated as Page's test (a.k.a. CUSUM). The joint use renders quicker detection and more overall sensitivity.

2. PROBLEM STATEMENT

Manufacturing abnormalities appear, for instance, in papermaking and cast metal surfaces. There exist imaging solutions to track down such process-related defects. The system grazed here uses a linear array of charge-coupled device (CCD) detectors in which the detectors scan their field-of-views in a direction orthogonal to the moving surface. For every performed scan, a linescan spatial surface profile is captured in each detector. The spatial extent of a defect can range arbitrarily, which restricts the use of off-line approaches in terms of undefined memory usage and processing time. In other words, a complete two-dimensional image could be provided, if restrictions concerning memory and computing capacity were ignored. Hence, the change detection methodology provides preferential tools for linescan outlier detection. Specifically, the arriving data needs to be processed line-by-line, i.e., sequentially. The recognition problem is not addressed here.

Desired inspection properties include few false alarms and missed detections, and low detection delay. Next, a new test is proposed and the discussed methods are outlined. The reader is encouraged to refer to [5, 7]. In further discussion, its application is suggested and some CCD detector responses for real measured linescan images are examined. Finally, the conclusions are drawn.

Thanks to the National Technology Agency of Finland (TEKES) and Nokia Foundation for funding and support. Special thanks to Mr. J. Rauhamaa, ABB Oy, for co-operation and the provided test images.
Author's email: juho.vihonen@tut.fi

3. 1-D PAGE-LIKE RECURSION FOR ON-LINE HYPOTHESIS TESTING

Page suggested the use of repeated test for the hypotheses:

$$H_0: \quad \theta = \theta_0 \tag{1}$$

$$H_1: \quad \theta = \theta_1. \tag{2}$$

The algorithm is based on the concept of the logarithm of the likelihood defined by

$$g(x) = \ln \frac{p_{\theta_1}(x)}{p_{\theta_0}(x)}, \tag{3}$$

where p_{θ_0} and p_{θ_1} represent the probability density functions of the random variables under two distributions. In the particular case where the distribution is Gaussian with mean value μ and constant variance σ^2 the statistic can be written as

$$g(x) = \frac{1}{\sigma}\Big(x - \frac{\mu_0 + \mu_1}{2}\Big). \tag{4}$$

Let k denote the time index. The proposed recursion augments the standard test and is given by

$$\arg\min_k \{S_k \geq h\}, \tag{5}$$

where

$$S_k = \max\{0, S_{k-1} + g(y_k + \epsilon_k \gamma)\}. \tag{6}$$

The point here is that generally faults range in duration and magnitude. Especially, short duration raises the risk of miss-detection for a long-term test, which is why the term $\epsilon_k \gamma$ is introduced in (6). Assume that ϵ_k is given by some auxiliary process observer:

$$\epsilon_k = \begin{cases} 1, & \text{likely transient state} \\ 0, & \text{normal state} \end{cases}. \tag{7}$$

By further setting

$$\gamma = \frac{\mu_1 - \mu_0}{2}, \tag{8}$$

the functional property of (4) reduces the recursion to minimum possible jump detector, whatever their magnitude [7]. The step (8) is simply for allowing unconstrained statistic development under Gaussian statistics. Partly steered by some auxiliary supervision, detection is declared when the statistic S_k exceeds h.

4. 2-D SEQUENTIAL LINESCAN PROCESSING

Let the jth observation of kth scan be denoted by y_k^j. The recursion for line-by-line sequential processing is the same as in Sect. 3, except that the expression (6) is replaced with

$$S_k^j = \max\{0, S_{k-1}^j + g(y_k^j + \epsilon_k^j \gamma) + \sum_{\substack{i \\ i \neq j}} \epsilon_k^j g(y_k^i + \epsilon_k^i \gamma)\}, \tag{9}$$

which continues with the score function accumulation according to the spatial extent of a transient. Suppose that two proposed algorithms are run in parallel; the first for detecting an increase in mean, and the second for detecting a decrease in mean. For example for positive tail of the distribution, accumulation is easily accomplished by

$$S_k^j = \begin{cases} \max\{\epsilon_k^n S_k^n\}, & \text{if } \epsilon_k^j = 1 \\ S_k^j, & \text{otherwise} \end{cases} \tag{10}$$

for all $n \in \{i | y_k^i > \mu_0\}$. So, the maximization of statistic value is acceptable only if the observation is in the right tail of the underlying distribution and has acceptable risk level.

5. MODEL AND RESIDUAL UTILIZATION

The value of ϵ_k^j may be based upon thresholding the current prediction error after an inverse filtering procedure, as it will be discussed next. The KF and the state-space model have been used in a variety of signal processing and control applications. Whereas the KF provides an elegant solution for residual generation, the surface modeling issue is highly complex. The general formulations of the chosen concepts are briefly summarized next.

5.1. Used surface model

Under a stable process state, the observed noise and surface responses are assumed to be realizations of pth-order Gauss-Markov process, i.e.,

$$s_k = -\sum_{i=1}^{p} a_i s_{k-i} + u_k, \tag{11}$$

where $u_k \sim N(0, \sigma_u^2)$. The statement may be regarded as an autoregressive, AR(p), process excited by white Gaussian noise [5] having state-space representation:

$$\mathbf{s}_k = \mathbf{A}\mathbf{s}_{k-1} + \mathbf{B}\mathbf{u}_k, \quad k \geq 0, \tag{12}$$

and $\mathbf{u}_k \sim N(0, \mathbf{Q}_k)$. An extensive treatment of AR-identification can be found in [8].

5.2. Residual utilization

Using the measurement model

$$y_k = \mathbf{h}^T \mathbf{s}_k + w_k, \tag{13}$$

$w_k \sim N(0, \sigma_w^2)$, and without specifying further details, the KF recursion can be written

$$
\begin{aligned}
\hat{\mathbf{s}}_{k|k-1} &= \mathbf{A}\hat{\mathbf{s}}_{k-1|k-1} && (14)\\
\mathbf{M}_{k|k-1} &= \mathbf{A}\mathbf{M}_{k-1|k-1}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T && (15)\\
\mathbf{k}_k &= \frac{\mathbf{M}_{k|k-1}\mathbf{h}}{\sigma_w^2 + \mathbf{h}^T\mathbf{M}_{k|k-1}\mathbf{h}} && (16)\\
r_k &= y_k - \mathbf{h}^T\hat{\mathbf{s}}_{k|k-1} && (17)\\
\hat{\mathbf{s}}_{k|k} &= \hat{\mathbf{s}}_{k|k-1} + \mathbf{k}_k r_k && (18)\\
\mathbf{M}_{k|k} &= (\mathbf{I} - \mathbf{k}_k\mathbf{h}^T)\mathbf{M}_{n|n-1}. && (19)
\end{aligned}
$$

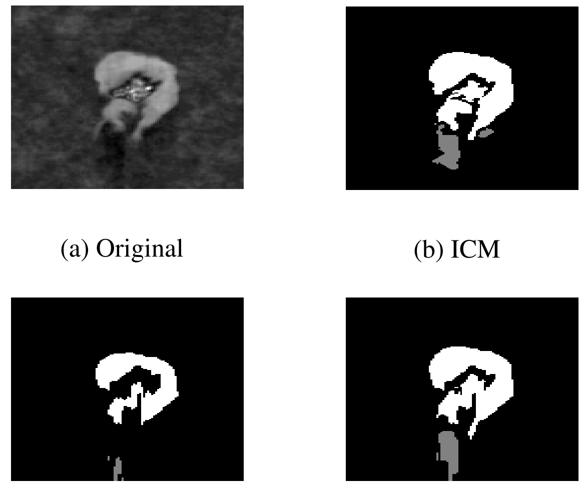
The optimality of the above recursive algorithm is based on knowing the quantities $\mathbf{Q}$ and σ_w^2. Note that the steady-state values can be precomputed by iterating. The KF residual monitoring is used for model validating. Based on the assumption that no failures have occured, the residual sequence is theoretically well described as a zero mean Gaussian sequence with variance $\sigma^2 = \sigma_w^2 + \mathbf{h}^T\mathbf{M}\mathbf{h}$.

The confidence bounds for ϵ_k^j are easily constructed. If $|r_k^j|$ is greater than 3σ, the current sample is declared as unacceptable, and the KF is initialized with the expected normal-state signal mean value. Each scan is filtered from both ends and minimum jump processing is used only if both thresholded residual sequences indicate fault. See [7] for more sophisticated analysis methods.

6. EXAMPLES

AR(4)-model is used with the recursive-in-order Burg identification for readily transient-free data. The stochastic noise input to the AR-model should represent the true noise of the process, which determines σ_u^2. σ_w^2 is set here to $0.0025\sigma_u^2$, which emphasizes the uncertainty of the model, but allows process tracking. The threshold h was chosen to provide false-alarm free performance, and $\mu_1 = \mu_0 + 4$.

Comparison is made between the proposed recursive scheme and the iterated conditional modes (ICM) [9], which is an iterative, well-known, and well-performing segmentation algorithm. The ICM was evaluated in the context of hypothesis testing. In Fig. 1 (a), a heavy pulp spot is captured. Badly wrinkled surface is observed in Fig. 2 (a). The original images are scaled for better view. Additional figures represent the image processed using different methods, and gray and white colors represent the different tails of the background distribution. The double-sided ICM iterated the images several times using 3x3 window with threshold and neighborhood influence parameters set to be the best found compromise for the both cases. The used window in (9) was chosen correspondingly, i.e., $i \in [j-1, j+1]$.



(a) Original (b) ICM

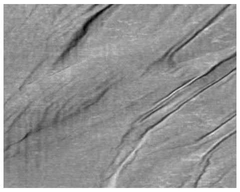
(c) Standard Page's test (d) Proposed Page-like test

Fig. 1. An example of a typical heavy pulp spot and its decomposition with different methods. Note that the detection is hardly delayed in (d), if compared to (b). Likewise, the performance is nearly equivalent. The gray area in (c) and (d) shows also the advantage of minimum jump detection.

7. DISCUSSION

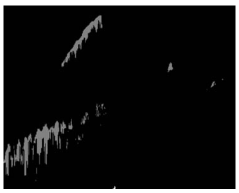
In the general case, the extension in (9) should treat a transient starting from y_k^{j-m} to its other edge y_k^{j+n} as one. It is not trivial to determine the true m and n of a transient, although Page's test could be used. Here, these parameters are chosen to be $m = n = 1$, which is obviously *ad hoc*. The used choice amounts to maximize the decomposition resolution. However, the implemented ICM was expected to give the ground truth in Figs. 1, 2. Even so, the proposed recursion seems to exhibit much alike results. In fact, the suggested method may be interpreted as a recursive version of the ICM. Note that the local unfit statistic is forced down and vice versa, which stems from (7). Hence, the phase (7) may be compared to the ICM initialization and the extension in (9) similarly to the conditional probability maximization.

The KF is designed to filter any deviations in the measurements and predictions using updates. Thus, the softer the change is, the harder the task becomes to tackle with residuals. The KF's error covariance becomes also easily large after identification. Hence, even a relatively large residual falls well within its bounds (not to even mention bounds with few step prediction using forward projected covariance). The used simple model is most suitable for cases where the process change is abrupt. This, on the other hand, is related to numerous factors such as pixel size, sweep time, and focusing. Hence, the possibility of using multiple models can't be ruled out.



(a) Original

(b) ICM



(c) Standard Page's test

(d) Proposed Page-like test

Fig. 2. An example of a very wrinkled paper surface and decomposition results. The ICM and the proposed recursion show both decent performances.

Keeping in mind the ICM's fixed window size, the recursive statistic incorporates the measurement time history. This gives the Page-like recursion an obvious advantage. The simple hypothesis testing justifies also the minimum jump processing, since it is unnecessary to test a sample twice. Unfortunately, the problem statement, to be precise, line-by-line processing results in lack of symmetry of the sequential testing. Note that a horizontally-oriented transient (Fig. 2) represents the pathological case for the standard Page's test, but the proposed recursion handles it pretty well.

8. CONCLUSIONS

We proposed new Page-like, minimum memory recursion with supervision to enhance and accelerate short-duration signal detection. It involves Kalman filter residual analysis with affordable computation. By recognizing that the Page's test is optimal for only some jump magnitude, the residual treatment can be used to liberate the score statistic development in clear outlier cases. Despite the *ad hoc* nature of the neighborhood use, which led to a recursive algorithm with features similar to the ICM, the same principle was adopted to linescan data processing with satisfactory outcome. The approach may be further improved by using change time estimation. Now the spatial alarm time from the Page's test is used for segmentation.

9. REFERENCES

[1] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Processing*, vol. 14, pp. 294–307, 2005.

[2] E. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.

[3] A. Wald, *Sequential Analysis*, Wiley, New York, 1947.

[4] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Statist.*, vol. 19, pp. 326–339, 1948.

[5] S. M. Kay, *Fundamentals of statistical signal processing, Estimation theory, Vol. 1*, Prentice-Hall, New Jersey, 1993.

[6] P. D. Hanlon and P. S. Maybeck, "Characterization of Kalman filter residuals in the presence of mismodelling," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, pp. 114–131, 2000.

[7] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes - Theory and Application*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[8] P. Stoica and R. Moses, *Introduction to spectral analysis*, Prentice-Hall, New Jersey, 1997.

[9] J. E. Besag, "On the statistical analysis of dirty pictures," *J. Royal Statist. Soc., Series B*, vol. 48, pp. 259–302, 1986.