# A NOVEL METHOD FOR QUICK ON-LINE SEGMENTATION BASED ON SPARSITY

*J. Vihonen*[⋆] *J. Rauhamaa*[†] *T. Huotilainen*[†] *A. Visa*[⋆]

[⋆] Tampere University of Technology, Department of Signal Processing, FI-33101 Tampere, Finland
[†]ABB Oy, Process Industry, Web Imaging Systems, FI-00381 Helsinki, Finland

## ABSTRACT

We propose a novel multi-estimate dynamic programming (DP) method for on-line detection and segmentation of significant anomalies in a video sequence. The method is based on the concept of sparsity, which means that we reduce visual features of each frame to a set of keypoints. In our line-scan application this is done by extracting only the intensity extrema. This way, we can decrease DP's inherent noise-amplifying tendency when building up the estimates of an anomaly. For detection improving, we introduce weights that express similarity between the spatial distribution of pixels forming so-called DP score sums and a reference representing their assumed distribution. The spatial dynamics estimation is improved by 30 % if compared to the intensity-only DP. Some 59 % change point recovery rate is attained in a web imaging application where illumination varies, contrasts are small, and the decision making time is limited to fractions of a second due to high-speed running of the web.

***Index Terms***— Estimation, change point segmentation.

## 1. INTRODUCTION

Matched filter has been widely used for detection of changes with spatial and temporal dynamics in imaging due to its optimality in maximizing the signal to noise ratio in the presence of additive random noise. However, because perfect matching is unlikely in reality, estimation based on dynamic programming (DP) has also been a subject of studies in the past.

Given a change dynamics model, the DP-based estimation is a process producing a multitude of potentially infinite sequences of pixels from the frames continuously observed. The ability to detect a change from these sequences is of a key interest. In this context, we consider a sparse representation of significant visual features and study a novel DP-based weighting method for sequential change point segmentation of them. In our approach, the sparsity is realized by extracting local maxima (minima) of each input frame in a prefiltering step. The weights express similarity between the spatial distribution of the pixels, that build up so-called DP score sums, and a reference based on explicit or implicit assumptions about the true distribution. In this way, this paper presents incremental progress on DP-based dynamics estimation over our previous work [1] where, along sequential run-length analysis, an algorithmic speed-gain by extracting extrema for the standard DP analyzed in detail in [2] was discovered.

The paper is organized as follows. The observation model, a review of the DP-based estimation, and words about cumulative sums of random variables from the viewpoint of change points are given in Sect. 2. Sect. 3 presents the motives of the work by first making a short note why the DP analyzed in [2] is prone to errors. Then technical operating characteristic computing in the sequential context is presented, since the ability of sequential tests to detect small changes and change points in possibly long samples of data is often superior if compared with fixed sample size tests or hard thresholding, see e.g. [3]. Sect. 4 contains simulation and experimenting with a multi-estimate version of the original simple DP, whose estimates are sequentially monitored. Conclusions are drawn in Sects. 5 and 6.

## 2. OBSERVATION MODEL AND ESTIMATION

After removing deterministic background components and improving detection sensitivity by various filtering methods that reduce noise and enhance the actual change, point representation of the change intensity can be given by

$$z_{\mathbf{r}_i} = A + v, \tag{1}$$

where $A$ is the change magnitude, $\mathbf{r}_i = [x_i, y_i]^T$ is the x,y-position vector of a pixel at the frame $i$ where $^T$ is the transpose operation, and $v$ is white Gaussian zero-centered noise with a constant variance. The x,y-position $\mathbf{r}_i$ is slowly varying according to a known model between successive frames in an appropriate imaging configuration. The magnitude $A$ may come from a particular distribution but we assume henceforth that its mean value is simply non-zero. In the matched filtering, the pixels the change (1) most likely occupies are found by

$$[\hat{\mathbf{r}}_i, \ldots, \hat{\mathbf{r}}_{i-N}] = \underset{\mathbf{r}_i^0, \ldots, \mathbf{r}_{i-N}^0}{\arg\max}\ \mathbf{I}_i \tag{2}$$

where the superscript "0" represents a known relationship or constraint between the successive position vectors of the change, and $\mathbf{I}_i = \{I(\mathbf{r}_i, \ldots, \mathbf{r}_{i-N})\}$ denotes the set of all possible (sufficient) statistics computed with the valid combinations of the pixels in the $N+1$ most recent frames at a time. The subscript $i$ refers to the most recently captured frame, termed also the "end-frame" later.

### 2.1. DP-based change estimation for imaging

If the x,y-coordinates of a change (1) vary in a random manner but within some physics-related limit between two successive frames, we can again base our estimate on the arguments of (2) with the difference that now the number of pixel combinations—out of which only one constitutes the change—grows exponentially as more frames are captured. A standard approach making feasible computation possible is to sum up intensity values of pixels among regions of a fixed size recursively for each end-frame pixel and then take the maximum as in (2); i.e., we can write

$$I_{\mathbf{r}_i} = \max_{\mathbf{r}_{i-1} \in R(\mathbf{r}_i)} \{I_{\mathbf{r}_{i-1}}\} + z_{\mathbf{r}_i} \tag{3}$$

for an end-frame pixel at $\mathbf{r}_i$ where $I_{\mathbf{r}_0} = 0$ for the initial condition and $R(\cdot)$ defines the set of plausible positions of the change at the next/previous frame with respect to an end-frame coordinate $\mathbf{r}_i$.[1] Due to the continuous flow of frames, the "oldest" pixel of those that build up the largest sum of $\{I_{\mathbf{r}_{i-1}}\}$ in (3) yields the estimate, which is thereby removed from the temporary sum of $N$+1 intensity values (not shown). Formally for $j = i, \ldots, i-N-1$ for all the sums $\mathbf{I}_j = \{I_{\mathbf{r}_j}\}$

$$\hat{\mathbf{r}}_{j-1} = \arg \max_{\mathbf{r}_{j-1} \in R(\hat{\mathbf{r}}_j)} \mathbf{I}_j \tag{4}$$

produces coordinates $[\hat{\mathbf{r}}_{i-1}, \ldots, \hat{\mathbf{r}}_{i-N-1}]$ of the pixels that build up the largest sum in (3) as in (2). Undoing the maximization by working backward is historically called backtracking. Numerous authors have studied the described and alike DP approaches during the past few decades; see e.g. [2, 4, 5] and the references therein.

Note that independence in a set of scores (3) is achieved only if the score sums contain no joint pixels, but this would constitute a severe restriction for building up the sums. Moreover, this is also a major reason why performance evaluation tools for the DP above, both from the viewpoint of detection and dynamics estimation, tend to lack accuracy.

### 2.2. Change point detection in time series

Consider a process which produces a potentially infinite sequence of observations $y_1, y_2, \ldots, y_i, \ldots$ of a random variable $y$. Sequential change point detection in its classic form denotes repeated testing of the two simple hypothesis:

$$\begin{aligned} \mathrm{H}_0 \quad &: \quad \theta = \theta_0 \\ \mathrm{H}_1 \quad &: \quad \theta = \theta_1. \end{aligned}$$

Before the unknown change time $t_0$, the parameter $\theta$ is equal to $\theta_0$ , and after the change it is equal to $\theta_1 \neq \theta_0$. The problem is then to detect the change in the parameter as quickly as possible. If we restrict the analysis to a single sequence of point-like values $y_1, y_2, \ldots, y_i, \ldots$ extracted by DP frame-by-frame, the detection of the change with a minimum number of observations can be described by the alarm time

$$t_a = \arg \min_i \{S_i \geq h\} \tag{5}$$

where $h$ is a threshold and $\{S_i; i \geq 0, S_0 = 0\}$ is a stochastic process with independent increments. The optimum cumulative sum (CUSUM) statistic is given by

$$S_i = \max\{0, S_{i-1} + g(y_i)\} \tag{6}$$

in which $g(y) = p_1(y)/p_0(y)$ with the known probability density functions $p_0(y)$ for noise-only and and $p_1(y)$ for change plus noise states; see e.g. [6, 7, 8]. The optimum theory also states that the best estimate of the change time $t_0$ is given by a renewal process of (6) wherein each renewal takes place when $S_i$ returns to 0. If the change is temporary, the switch from $\theta_1$ back to $\theta_0$ can be described by

$$t_b = \arg \min_i \{m_i - S_i \geq \delta\} \tag{7}$$

where $0 \leq \delta < h$ and

$$m_i = \max_{k \leq i} \{S_k\} \tag{8}$$

is the current maximum value after the last renewal of (6), which yields the estimate of the switch time $t_1$. The estimates of the change points $t_0$ and $t_1$ segment the change in time.

For the statistic, we will use

$$g(y) = y - \nu, \tag{9}$$

as various criteria like the above can lead to it, see e.g. [9], and because the so-called run-length distribution of (5) can be presented for standardized parameters $\mu - \nu$ and $h$, see e.g. [10]. Here $\mu$ is the mean of $y$, and $\nu$ denotes a minimum interesting magnitude of a jump. Note the relation of (6), (9), and (3).

## 3. STATISTICAL REASONING

The statistical properties of (3) can be studied by reverting to the maximization process. Given that (4) renders a single "best" cumulant out of i.i.d. $z_1 < z_2 < \ldots < z_j$ realizations of a random variable $z$ denoting intensity, the probability density could be given by

$$f_\jmath(z) = \jmath \, [F(z)]^{\jmath-1} f(z), \tag{10}$$

where $F$ is the cumulative distribution and $f$ is the probability density function of $z$. Fig. 1 plots the densities for a few different numbers of plausible coordinates $\jmath$ defined by $R$ under Gaussian statistics. By the optimality principle it would be sufficient to retain the estimate with the highest score only. However, as illustrated, the noise becomes greatly amplified, which one may expect to gain strength from score dependencies when working with multiple frames in practice. This motivates reconsidering (3).

[1] If we know that a change is positioned at a pixel at some coordinate $\mathbf{r}_j = [x_0, y_0]^T$, then the set of plausible positions e.g. at the previous frame can be defined as $R(\mathbf{r}_j) = \{\mathbf{r}_{j-1} | x, y \in \mathbb{Z}, (x-x_0)^2 + (y-y_0)^2 \leq r^2\}$, where $\mathbb{Z}$ denotes the set of integers.

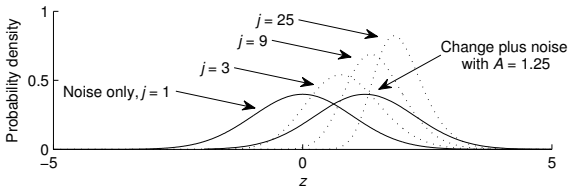



**Fig. 1**. Noise amplification process as the DP algorithm's search region pixel count is increased. $f_\jmath(z)$ denotes the probability density function of z as given in (10). For comparison, the density of the change is also illustrated in the same graph.

### 3.1. Weights based on spatial change dynamics

If a reference of the dynamics of the change would be available, corrupted by noise or not, one could realize the matching of the reference coordinates, forming a vector $\mathbf{C}_{\text{ref}}$, and coordinates of the pixels that build up a score sum in (4), stacked in a vector $\mathbf{C}_{\mathbf{r}_i}$, by using for example the $l^2$-norm as a measure of dissimilarity. Then the $l^2$-norm squared could be calculated as

$$d(\mathbf{C}_{\text{ref}}, \mathbf{C}_{\mathbf{r}_i}) = \frac{||\mathbf{C}_{\text{ref}} - \mathbf{C}_{\mathbf{r}_i}||_2^2}{N}, \tag{11}$$

which is a non-central chi-squared random variable.[2] Based on this, we define the weight

$$w(\mathbf{C}_{\text{ref}}, \mathbf{C}_{\mathbf{r}_i}) = \frac{1}{1 + d(\mathbf{C}_{\text{ref}}, \mathbf{C}_{\mathbf{r}_i})} \tag{12}$$

that is to be used in a manner illustrated in Fig. 2 to improve detectability. While the use of a large number of coordinates counters noise, weights close to unity can be attained only if the reference contains no systematic errors.

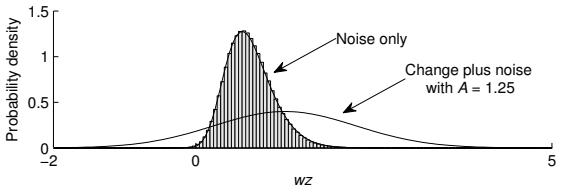



**Fig. 2**. When multiplied by (12), the density of the observations drawn with $\jmath = 9$ in Fig. 1 is here shifted to the left ($\sigma^2_{\text{ref}} + \sigma^2_{\mathbf{r}_i} = 2$, $d(\mathbf{Y}_{\text{ref}}, \mathbf{Y}_{\mathbf{r}_i}) = 0$). The probability density of the change is the same as in Fig. 1.

### 3.2. Detectability improved

When the densities overlap like above, a number of samples depending upon the underlying statistics is required for discrimination. The receiver operating characteristic (ROC) curves shown in Fig. 3 are the cumulated values of run-length probability mass distribution for (5) computed using the noise-only for $\jmath = 9$ and change plus noise densities in Figs. 1 and 2. The curves are produced for $h = 5$ (using 250 Gauss-Legendre points, see [10]) so that the parameter $\nu$ is chosen to produce a fixed mean false-alarm interval of about 100 observations in the both a) and b), $P_{\text{d}}$ being the probability of detection and $P_{\text{fa}}$ being the probability of false alarm as usual. These curves are suggestive due to the fact that the analysis is one-dimensional, but one could apply expressions derived in [1] for a ROC generalization to an imaging setup by using the run-length probability mass distributions produced for Fig. 3. Here we omit this since $P_{\text{d}} \gg P_{\text{fa}}$ just after 10 cumulant, showing that assigning the spatial domain weights can provide the detectability needed.

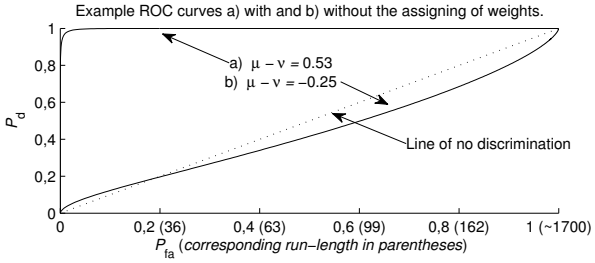



**Fig. 3**. ROC curves with and without the weights using the densities in Figs. 1 and 2, $\jmath = 9$. While a) indicates quick detection, the probability of false alarm is very high in b).

## 4. MULTI-ESTIMATE CHANGE SEGMENTATION

The noise amplifying nature of the maximization step in (3) is intrinsic to the parameters of the method and is not due to improper use of DP, which motivates developing the simple DP further. In particular, one needs to reduce the typical noise-amplifying ambiguity embedded in the DP's search space in order to attain efficiency like in Fig. 3. Explaining the term "multi-estimate", we shall write

$$\mathbf{I}_{\mathbf{r}_i} = \{\text{prune}(\mathbf{I}_{i-1}) + z_{\mathbf{r}_i}\} \tag{13}$$

for an end-frame pixel at $\mathbf{r}_i$ to explicitly include a more DP-style pruning of all the cumulated score sums up to and at the previous frame $\mathbf{I}_{i-1} = \{I_{\mathbf{r}_{i-1}} | \mathbf{r}_{i-1} \in R(\mathbf{r}_i)\}$ into our notation. Then employing the concept of sparsity, the following steps are taken:

- For prefiltering, we suggest extracting local maxima (minima) framewise. Only these brightest (darkest) pixels, visual keypoints, will be used in the estimation.
- For pruning, we suggest retaining the largest score sums out of those sums in $\mathbf{I}_{i-1}$ having one or more pixels in common. This is simple to implement since it does not depend on the change's (1) actual dynamics.

The first step makes it possible to uncover the spatial dynamics of a low contrast change while the second limits computational burden. For the theoretically quickest detection (5) followed by the change point segmentation (7), we compute

$$\mathbf{S}_{\mathbf{r}_i} = \{\mathbf{S}_{i-1} + g(w_{\mathbf{r}_i} z_{\mathbf{r}_i})\}^+ \tag{14}$$

[2] Given that the both $\mathbf{C}_{\text{ref}}$ and $\mathbf{C}_{\mathbf{r}_i}$ contain both deterministic components $\mathbf{Y}_{\text{ref}}$ and $\mathbf{Y}_{\mathbf{r}_i}$ plus zero-centered i.i.d. Gaussian noise with variances $\sigma^2_{\text{ref}}$ and $\sigma^2_{\mathbf{r}_i}$, the mean is $\mathbf{E}(d(\mathbf{C}_{\text{ref}}, \mathbf{C}_{\mathbf{r}_i})) = d(\mathbf{Y}_{\text{ref}}, \mathbf{Y}_{\mathbf{r}_i}) + \sigma^2_{\text{ref}} + \sigma^2_{\mathbf{r}_i}$ and variance $var(d(\mathbf{C}_{\text{ref}}, \mathbf{C}_{\mathbf{r}_i})) = \frac{4(\sigma^2_{\text{ref}} + \sigma^2_{\mathbf{r}_i})^2 + 4(\sigma^2_{\text{ref}} + \sigma^2_{\mathbf{r}_i})^2 d(\mathbf{Y}_{\text{ref}}, \mathbf{Y}_{\mathbf{r}_i})}{N}$.

so that $\mathbf{S}_{i-1} = \{S_{\mathbf{r}_{i-1}} | \mathbf{r}_{i-1} \in R(\mathbf{r}_i)\}$ using the pixels that simultaneously build up the DP score sums (13) in the manner discussed above. The plus sign denotes the same renewal process shown in (6) wherein each renewal takes place when a statistic returns to 0, i.e. $\mathbf{S}_{\mathbf{r}_i} = \{\max(0, S_{\mathbf{r}_i})\}$. A high enough statistic implies the presence of a change.

The two-stage pruning above can result in high computational demands, for example easily eating away the speed-gain we found in [1], but the complexity is scalable. Also the possibility of parallel computing exists. We have controlled the total number of the DP score sums (13) per frame and their redundancy by frequency domain filtering of the frames, but here we omit such practicalities of low noise operation.

### 4.1. Simulated estimation error experiment

As a closure to results in the Figs. 1-3 for $\jmath = 9$, Fig. 4 provides a simulated example of estimation error of the multi-estimate DP assuming that a change (1) is present between 100 successive line frames.[3] For analytical simplicity, the change dynamics are linear of the first degree (a constant shift in position $\mathbf{r}_i$ equal to one pixel per frame) in the image plane which, without better knowledge or precomputed reference lookup tables, is here and in the following least-squares estimated for (11). A total of 20 frames is used at a time, and an equivalent matched filter (2) operates as the reference.

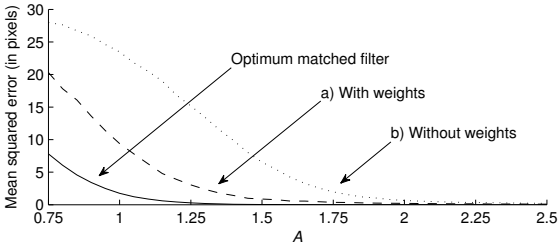



**Fig. 4**. Error curves where a) approaches the optimum at much lower signal levels than b), the simple DP-based estimation. As real changes' image plane spatial dynamics are often unique, optimality shown here is hardly ever established.

The relation of the above to prior work is that, for the given model (1) and details in Sect. 2, the simple DP has a fundamental performance limit at $A < 2$, which is depicted in [2]. The limit, also observable in Fig. 4 b), becomes practically invariant to the number of frames $N$ used by the DP and is due to the noise-amplifying nature of (3) and (4) like shown previously. Here, as only the local maxima (minima) are considered via prefiltering of the Gaussian distributed data, some 33 % of pixels statistically *likely* to contain a point-like positive (negative) change per frame make the sparse set of keypoints that updates the estimates (13). The reduction of data makes their spatial domain dynamics matchable to the reference distribution of pixels as in (12), showing some 30 % average improvement of the estimation error between a) and b) above. However, the theoretically achievable improvement is only limited by the statistics of (11) as detailed in Sect. 3.1. Since the scores in (13) follow an extreme value distribution, see e.g. [5], it is also elementary to see whether or not computing a weight is worth the effort in a more general or higher-order setup (i.e. if $A \gg 2$).

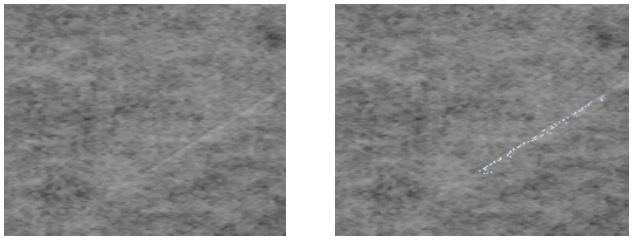
[3] When the 100th frame is processed, an estimate in a maximum likelihood sense is obtained by backtracking all the frames with and without assigning the weights using (4). This reflects the performance when change points are accurately on-line estimated with (13) and (14).

### 4.2. Change point segmentation experiment

Fig. 5 a) presents a low-contrast surface anomaly with spatial distribution somewhat linear of the first order. The vertical extent of the image is obtained by gathering sequentially rows of horizontally oriented line frames with a high speed line-scan CCD camera at a laboratory-level paper mill setup. Note the qualitative difference between a change in the mean value of the signal, as defined in (1), and a change in the behavior around a mean level, as shown here. Since each anomaly is unique, effectiveness is achieved by

$$w(\mathbf{C}_{\text{ref}}, \mathbf{C}_{\mathbf{r}_i}) = \begin{cases} 1 & \text{if } d(\mathbf{C}_{\text{ref}}, \mathbf{C}_{\mathbf{r}_i}) < \tau \\ \frac{1}{1+d(\mathbf{C}_{\text{ref}}, \mathbf{C}_{\mathbf{r}_i})} & \text{otherwise} \end{cases} \quad (15)$$

where is $\tau$ is the maximum distance selected by deterministic speculations about the likely difference (11) in practice, discussed also in the footnote[2]. As the anomaly extends over numerous line frames and is not separable by intensity only, the sequential segmentation with (13) and (14) is used in b) showing quite accurately estimated change points out of all the 8-bit gray-scale values. The margin between the noise-only and change plus noise statistics (14) was about ten fold.



a) b)

**Fig. 5**. Low-contrast anomaly in a) and multi-estimate segmentation result in b). Here $\tau = 2.7$, $N = 20$ to counter higher frequency fluctuation. The end-frame maxima were connected to the scores (13) within the 17 nearest pixels.

Results of a subjective segmentation quality test are presented in Table 1. The labels on top denote different types of paper grades ranging from drawing paper (i, 130 g/m$^2$) to standard white (ii, 80 g/m$^2$) and red-colored copy paper (iii, 80 g/m$^2$), thick (iv, 160 g/m$^2$) and coated color copy paper (v, 80 g/m$^2$), and calendered typewriter paper (vi, 70 g/m$^2$) selected to represent the wide variety of paper formation types.

All the samples contained a surface anomaly alike that shown in Fig. 5 a) and were viewed from different angles. For instance the result in b) was evaluated to be about 70 % correct from the viewpoint of the anomaly's length, but the absolute truth is not easily stated. Nonetheless, the ratings show that choosing a large-enough number of frames filters higher frequency non-linearities in the intensity levels, unachievable in a simple intensity only DP setup without additional measures.

**Table 1**. Segmentation quality test results using different paper grades captured at 42 $\mu$s scan rate with paper passing the camera 3.9 m/s and horizontal pixel resolution of 0.125 mm. The "finer" the paper grade, the better the recovery rate.

| Paper grade: | i | ii | iii | iv | v | vi |
|---|---|---|---|---|---|---|
| Recovery rate (%): | 37 | 44 | 44 | 67 | 79 | 83 |

## 5. DISCUSSION

Though the ratings with 59 % average in the Table 1 may seem low, the percentages are adequate for an anomaly recognition procedure, which is typically adopted when the problems of on-line detection and change segmentation coexist, to more holistic change characterization without stringent real-time requirements. Thereby undersegmentation is generally less critical. For fair comparison, the DP parameters were not paper grade specifically tuned. But even if they had been tuned, the optical fluctuation cannot be eliminated completely from partially masking low-contrast changes as shown in Fig. 5. Thus, some simple clustering, rank ordering at the pruning stage of alike DP scores, or spatial domain preprosessing as discussed next could come to mind.

The DP-based estimation can be applied to a variety of problems in which the target satisfies point representation or can be transformed into such. In remote sensing, for example, one refers to track-before-detect processing designed for clutter-obscured dim signal detection often comparable to that in Fig. 5. If preceded by enhancement filtering that is matched to the expected spatial frequency of a known anomaly, the DP may perform well without the "smarter" estimate build-up or spatial monitoring. That is, achieving the efficiency of detection in Fig. 3 and the accuracy of estimation close to that of the matched filter in Fig. 4 is linked to the question, what preprocessing provides best possible detection sensitivity? Since finding a solution to this problem often promotes simpler visual feature representation, our keypoint-based DP can be a viable tool due to the spatial information embedded in its estimates with or without the CUSUM-based segmentation.

## 6. CONCLUSION

We have studied a DP solution for multiframe on-line detection and segmentation of small low-contrast changes. Our multi-estimate DP uses visually significant pixels, viewed as keypoints, by selecting only the local intensity maxima (minima) framewise. Like the pruning presented, also the weights we introduced for detection and segmentation purposes make use of spatial distribution of the selected pixels. From the viewpoint of our line-scan application, picking up intensity extrema per frame is probably the simplest way to achieve a reduced, sparse, representation of visual features. Thus, we consider the simulated 30 % gain in accuracy and the empirical 59 % recovery rate of low-contrast change points in the experiments encouraging.

## 7. REFERENCES


[1] J. Vihonen, J. Jylhä, T. Ala-Kleemola, M. Ruotsalainen, J. Kauppila, T. Huotilainen, J. Rauhamaa, and A. Visa, "Directional filtering for sequential image analysis," *IEEE Sig. Process. Lett.*, vol. 15, pp. 902–905, 2008.

[2] S. Tonissen and R. Evans, "Performance of dynamic programming techniques for track-before-detect," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 4, pp. 1440–1451, Oct. 1996.

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

[4] J. Arnold, S. Shaw, and H. Pasternack, "Efficient target tracking using dynamic programming," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 44–56, Jan. 1993.

[5] L. Johnston and V. Krishnamurthy, "Performance analysis of a dynamic programming track before detect algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 1, pp. 228–242, Jan. 2002.

[6] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, 1971.

[7] G. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, 1986.

[8] Y. Ritov, "Decision theoretic optimality of the CUSUM procedure," *Ann. Statist.*, vol. 18, no. 3, pp. 1464–1469, 1990.

[9] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, Wiley, New York, NY, 2000.

[10] A. Luceño and J. Puig-Pey, "Evaluation of the run-length probability distribution for CUSUM charts: Assessing chart performance," *Technometrics*, vol. 42, no. 4, pp. 411–416, 2000.