ELSEVIER

Signal Processing: *Image Communication* 7 (1995) 147–160

SIGNAL PROCESSING: IMAGE COMMUNICATION

Bayesian algorithms for adaptive change detection in image sequences using Markov random fields[1]

Til Aach*, André Kaup

Institute for Communication Engineering, Aachen University of Technology (RWTH), Melatener Strasse 23, D-52056 Aachen, Germany

Received 28 March 1994

Abstract

In many conventional methods for change detection, the detections are carried out by comparing a test statistic, which is computed locally for each location on the image grid, with a global threshold. These 'nonadaptive' methods for change detection suffer from the dilemma of either causing many false alarms or missing considerable parts of non-stationary areas. This contribution presents a way out of this dilemma by viewing change detection as an inverse, ill-posed problem. As such, the problem can be solved using prior knowledge about typical properties of change masks. This reasoning leads to a Bayesian formulation of change detection, where the prior knowledge is brought to bear by appropriately specified a priori probabilities. Based on this approach, a new, adaptive algorithm for change detection is derived where the decision thresholds vary depending on context, thus improving detection performance substantially. The algorithm requires only a single raster scan per picture and increases the computional load only slightly in comparison to non-adaptive techniques.

Keywords: Image analysis; Image coding; Context-adaptive change detection; Markov random fields

1. Introduction

The detection and accurate localization of intensity changes between subsequent frames of image sequences is a crucial issue for coding of moving video [10, 21, 30], in particular for region-oriented techniques [13, 17, 20], as well as for a variety of tasks in image analysis [8, 14, 19]. In any one of these applications, the purpose of change detection is to locate moving objects in the image plane by exploiting the fast temporal grey level variations caused by moving objects. To separate fast changes from slow drifts in intensity, which may e.g. be due to varying scene illumination, a highpass filtering operation is often carried out by subtracting subsequent pictures of the image sequence to be processed.

An inherent difficulty when evaluating difference images is posed by the presence of noise, which gives rise to intensity changes covering moving areas as well as stationary ones. On the other hand, it is by no means certain that motion always causes perceptible temporal variations, as these are also dependent on the spatial grey level gradient

[1] Dedicated to Prof. Dr.-Ing. Hans Dieter Lüke on the occasion of his 60th birthday.

* Corresponding author. Present address: Philips GmbH Research Laboratories, Weisshausstrasse 2, D-52066 Aachen, Germany. Tel.: + 49 241 6003 567. Fax: + 49 241 6003 465.

SSDI 0923-5965(95)00003-8

(cf. [16]). Change detection by thresholding test functions computed from local samples of grey level differences thus suffers from the dilemma of either causing many false alarms or failing to detect considerable parts of genuinely moving areas [5].

The reason for the poor operating characteristic of this approach is that local samples do not contain any information about the global properties of moving regions [6].[1] Regions corresponding to moving objects tend to be of compact shape with smooth boundaries. Regions caused by false alarms almost never exhibit these properties, on the contrary, they manifest themselves in irregular speckles spread randomly over the image plane. The main objective of this paper is to integrate this kind of prior information into change detection algorithms in order to distinguish better between 'real' changes and noise-related detection errors, thus allowing substantial improvements in detection performance to be achieved. In practice, this leads to the decision on whether a picture element is to be labelled as 'changed' or 'unchanged' being made in context with other decisions regarding its neighbours (cf. [15]).

To find a formalism which allows the taking into account of prior knowledge about global region properties it is helpful to add change detection to the list of inverse problems of low level vision, which include edge detection, optic flow and surface reconstruction [7, 25, 26]. This reasoning leads to a Bayesian formulation for the problem of change detection, where our prior knowledge can be brought to bear by appropriately specified a priori probabilities. A tool well suited to the purpose of expressing our prior knowledge is formed by Gibbs/Markov random fields, which in the past have been used with considerable success in a variety of image segmentation tasks [2–4, 12]. Starting from this framework, we develop an adaptive change detection algorithm, which, due to its non-iterative nature, is very attractive also from a computational point of view. In the next section, we first formulate change detection as a Bayesian estimation problem, and specify its main components, viz. likelihood ratio and a priori probabilities. From this formulation, decision rules and a proposal for a practical implementation will be derived.

[1] This difficulty thus also affects change detection algorithms which do not directly evaluate grey level difference images, like [18, 23], since they suffer from the same problem of looking at local samples only.

2. Change detection as a Bayesian estimation problem

Let $D = \{d(k)\}$ denote the grey level difference image, with $d(k) = y_1(k) - y_2(k)$, where $y_1(k)$ and $y_2(k)$ are the grey levels at pixel location k of two subsequent pictures Y_1 and Y_2 of an image sequence. The change mask Q consists of a binary label $q(k)$ for each pixel k on the image grid. Each label $q(k)$ either takes the value $q(k) = u$ ('unchanged') if the observed grey level difference $d(k)$ supports the hypothesis that it is due to camera noise only (null hypothesis H_0), or the value $q(k) = c$ ('changed') if the observed value of $d(k)$ does not support this assumption (alternative hypothesis H_1). As a special case of a Bayesian estimate, we try to estimate the change mask Q such that its a posteriori probability $\Pr(Q \mid D)$ given the difference image D is maximized (MAP estimate).

Let us for the moment assume that the values of the labels $q(k)$ are known for all picture elements k except for one element i. Estimating Q then reduces to deciding between $q(i) = u$ and $q(i) = c$. The change mask resulting from $q(i) = u$ is denoted by Q_u^i, and that produced by $q(i) = c$ is termed Q_c^i. The decision rule can thus be written as

$$\frac{\Pr(Q_u^i \mid D)}{\Pr(Q_c^i \mid D)} \underset{c}{\overset{u}{\gtrless}} t, \tag{1}$$

with t being a decision threshold. This notation means that the outcome of the decision is $q(i) = u$ ('unchanged') if the left-hand side of (1) exceeds t, otherwise it is $q(i) = c$ ('changed'). For $t = 1$, this decision selects from the change masks Q_u^i and Q_c^i that one with highest a posteriori probability. Using Bayes' theorem, this decision rule can be rewritten as

$$\frac{p(D \mid Q_u^i)}{p(D \mid Q_c^i)} \underset{c}{\overset{u}{\gtrless}} t \frac{\Pr(Q_c^i)}{\Pr(Q_u^i)}, \tag{2}$$

where $p(D \mid Q)$ denotes the conditional probability density of the observed difference image D given a change mask Q, which acts as the likelihood function for Q. $\Pr(Q_u^i)$ and $\Pr(Q_c^i)$ are the a priori probabilities for Q_u^i and Q_c^i, respectively.

We now assume that the grey level differences $d(k)$ are conditionally independent, i.e. $p(D|Q) = \Pi_k\, p(d(k)|q(k))$. This assumption is certainly justified in unchanged areas, where the observed differences are regarded as being caused by camera noise only. Grey level differences in changed areas, however, are correlated [11], so that, at a first glance, they may not be assumed as independent. Nevertheless, we found that for the purpose of change detection, these statistical dependencies can in practice be neglected without perceptible deterioration in detection performance. Experimental evidence supporting this view is given in Appendix A. With this assumption, (2) can be simplified to

$$\frac{p(d(i)|H_0)}{p(d(i)|H_1)} \underset{c}{\overset{u}{\gtrless}} t\,\frac{\Pr(Q_c^i)}{\Pr(Q_u^i)}, \tag{3}$$

with $p(d(i)|\mathrm{H}_j)$ denoting the likelihoods for the hypotheses H_j, $j = 0, 1$, with respect to pixel i.

To make the detection algorithm more reliable, the decision to be taken should not be based on the grey level difference $d(i)$ at pixel i only, but on a local sample $\boldsymbol{d}_i$ comprising several differences $d(k)$ (see e.g. [5, p. 168]). The sample $\boldsymbol{d}_i$ is conveniently formed from the differences $d(k)$ lying inside a small sliding window w_i centred at location i. To incorporate the sample into the detection approach, (3) is slightly modified to

$$\frac{p(\boldsymbol{d}_i|\mathrm{H}_0)}{p(\boldsymbol{d}_i|\mathrm{H}_1)} \underset{c}{\overset{u}{\gtrless}} t\,\frac{\Pr(Q_c^i)}{\Pr(Q_u^i)}. \tag{4}$$

In contrast to (3), this rule decides on whether or not the null hypothesis can be accepted based on the entire sample $\boldsymbol{d}_i = \{d(k)|k \in w_i\}$.

To convert (4) into a practical decision rule, assumptions must be made for the conditional densities $p(d(k)|\mathrm{H}_j)$, $j = 0, 1$, as well as for the a priori probabilities. Commencing with $p(d(k)|\mathrm{H}_j)$, we assume the grey level differences to obey zero-mean Gaussian distributions with variances σ_0^2 and σ_1^2 for H_0 and H_1, respectively. As changed areas typically exhibit differences of large magnitude, the variance σ_1^2 is much larger than the variance σ_0^2 caused by noise; estimates yield $\sigma_1^2 > 100\sigma_0^2$. The above decision rule can now be rewritten as

$$\exp\left\{-\frac{1}{2}\left(1-\frac{\sigma_0^2}{\sigma_1^2}\right)\overline{\Delta_i^2}\right\} \underset{c}{\overset{u}{\gtrless}} t\left(\frac{\sigma_0}{\sigma_1}\right)^{N_w}\frac{\Pr(Q_c^i)}{\Pr(Q_u^i)}, \tag{5}$$

with $\overline{\Delta_i^2}$ being the normalized square sum of grey level differences $d(k)$ inside w_i, i.e.

$$\overline{\Delta_i^2} = \frac{1}{\sigma_0^2}\sum_{k\in w_i} d^2(k). \tag{6}$$

N_w denotes the size of the window w_i in picture elements.

As $\sigma_1^2 \gg \sigma_0^2$, the fraction σ_0^2/σ_1^2 may be dropped in (5). Taking the logarithm on both sides of (5) yields

$$\overline{\Delta_i^2} \underset{u}{\overset{c}{\gtrless}} \underbrace{-2\ln\left[t\left(\frac{\sigma_0}{\sigma_1}\right)^{N_w}\right]}_{t_s} + 2\ln\frac{\Pr(Q_u^i)}{\Pr(Q_c^i)}. \tag{7}$$

The decision threshold on the right-hand side of (7) consists of a fixed portion t_s, which is independent of Q_u^i and Q_c^i, and of a portion which depends on the logarithm of the a priori probabilities of the solutions. If Q_u^i has higher a priori probability, the logarithm is positive and raises the threshold, thus biasing the decision in favour of $q(i) = u$, as intended. Conversely, $\Pr(Q_u^i) < \Pr(Q_c^i)$ results in a decreased threshold, hence favouring $q(i) = c$.

2.1. Non-adaptive change detection

Let us suppose for the moment that we have no prior knowledge with respect to the expected change masks. We thus have no information about which one of the two possible change masks, Q_u^i or Q_c^i, has higher a priori probability. This can be expressed mathematically through $\Pr(Q_u^i) = \Pr(Q_c^i)$. Correspondingly, the rightmost logarithm of (7) vanishes, depriving the decision threshold of its adaptivity. What remains is the global decision threshold t_s. Instead of specifying t_s in terms of σ_0, σ_1 and t, as indicated in (7), it is more practical to

couple t_s to the rate α of false alarms associated with the test. As the normalized square sum $\overline{\Delta_i^2}$, given the null hypothesis H_0, is known to obey a χ^2 distribution with N_w degrees of freedom, the threshold t_s can be determined from

$$\Pr(\overline{\Delta_i^2} > t_s | H_0) = \alpha, \tag{8}$$

once an acceptable false alarm rate α has been chosen. This procedure is termed a significance test [28, 29, 5], with the false alarm rate α being called the significance. As no prior knowledge is brought to bear, the resulting non-adaptive technique is associated with the plight of either missing considerable parts of moving objects or producing many false alarms. Figs. 1–3 show original sequences, which are used as examples to demonstrate this behaviour in Figs. 4 and 5. For this experiment as well as for the following investigations, a window of size 5×5 pixels, i.e. $N_w = 25$, was utilized throughout.

The choice of an appropriate false alarm rate α is often not guided by mathematical considerations only, but depends also on the consequences the two

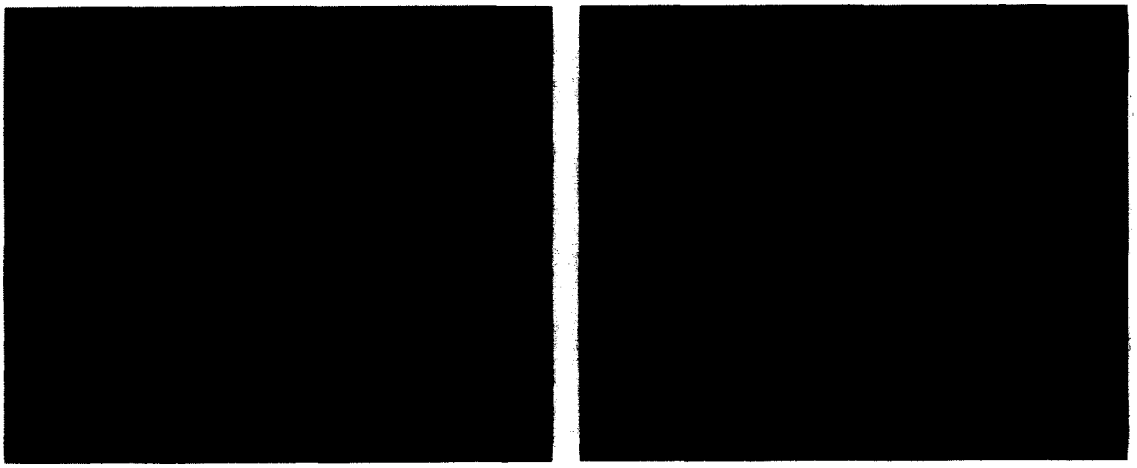

Fig. 1. First and third frame (256×256 pixels) of a speaker sequence.

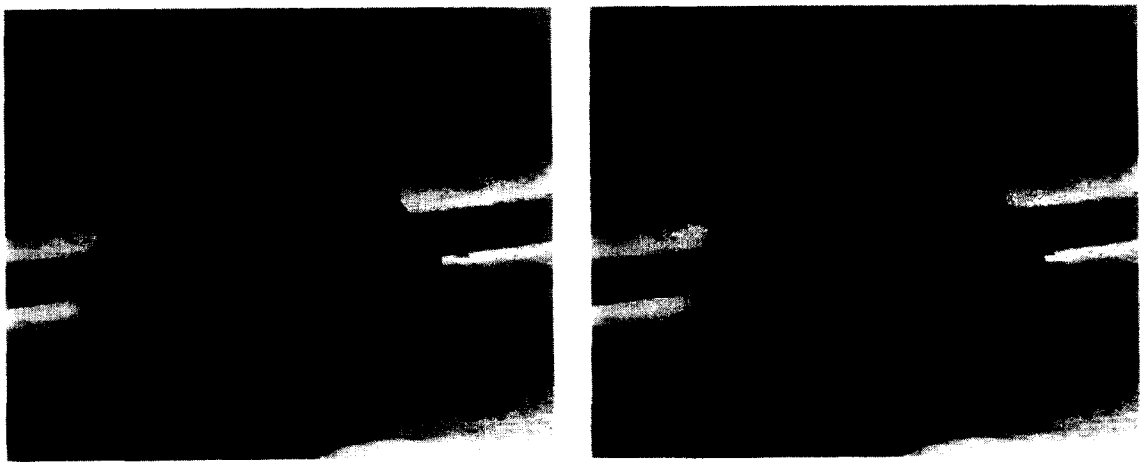

Fig. 2. Two subsequent frames (256×256 pixels) of a traffic scene.

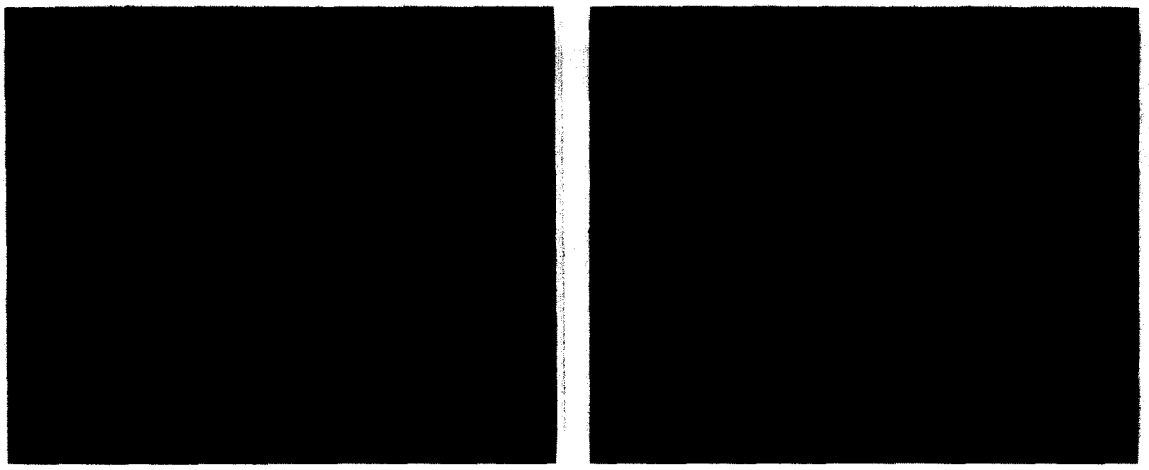

Fig. 3. Portions of size 256 × 256 pixels from frames no. 80 and 81 of the sequence *Miss America*.

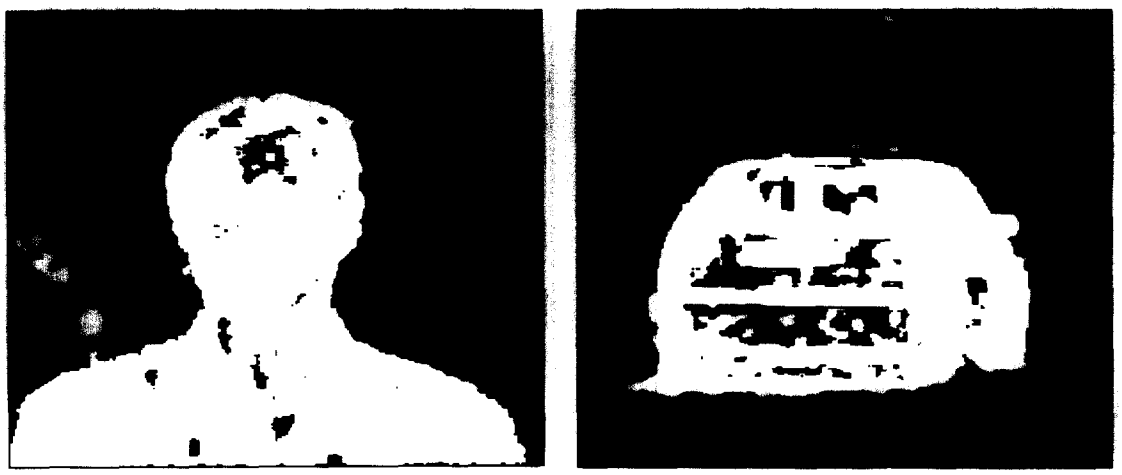

Fig. 4. Change masks obtained from Figs. 1 and 2 using the test statistic $\overline{\Delta_i^2}$, $\alpha = 10^{-6}$, $t_s = 74.5$. While the background is nearly error free, considerable portions of the moving objects could not be detected. (The blocky regions marked as changed in the background of the traffic scene are caused by pixel digitization errors which occurred during original image acquisition. These errors appear in difference images as differences of large magnitude.)

possible types of errors, false alarms and misses, incur. In image coding, false alarms cause an increased data rate, whereas a miss causes visible image quality degradations. From an image quality point of view, one would therefore prefer higher values for α, as in Fig. 5 (cf. also [5]).

Let us briefly note that employing Laplacian distributions to model the conditional densities $p(d(k)|\mathrm{H}_j)$ of the grey level differences, as proposed e.g. in [9], leads to a very similar framework (see [1, 5]). In this case, the local square sum $\overline{\Delta_i^2}$ is to be replaced by the absolute sum

$$\bar{\Delta}_i = \frac{2\sqrt{2}}{\sigma_0} \sum_{k \in w_i} |d(k)|, \tag{9}$$

which, given the null hypothesis, obeys a χ^2 distribution with $2N_w$ degrees of freedom. The

Fig. 5. Change mask obtained from Fig. 3 using the test statistic $\overline{\Delta_i^2}$, $\alpha = 10^{-2}$, $t_s = 44.3$. Due to the higher α, there occur numerous detection errors in the background.

performance of the detection algorithm remains nearly unaffected by the choice between $\overline{\Delta_i^2}$ and $\tilde{\Delta}_i$ (cf. also [31]).

2.2. *The a priori probability*

The change masks shown so far emphasize the already mentioned opposite properties of regions corresponding to moving objects and regions due to detection errors: while objects tend to manifest themselves in compact regions of preferably smooth shape, detection errors typically appear as small, scattered regions. By specifying the a priori probability such that smooth regions are more probable to occur than irregular ones, these properties can be exploited to improve the detection performance of change detectors in moving areas while simultaneously suppressing false alarms in stationary background. An expression well suited for the a priori probability can be found by describing the change masks as samples from two-dimensional Gibbs/Markov random fields. The a priori probability is then given by

$$\Pr(Q) = \frac{1}{Z}\exp\{-E(Q)\}, \tag{10}$$

with Z being a normalization constant. $E(Q)$ is a so-called energy term, which assesses the state of the change masks Q. In analogy to statistical physics where this model has its origin, Eq. (10) favours states of low energy. Consequently, $E(Q)$ should be specified such that the energy is low when the regions occurring in Q exhibit smooth boundaries, whereas irregular speckles should result in increased values for the energy.

The smoothness of region shapes can be evaluated by considering the so-called border pixel pairs that are associated with a change mask Q (see e.g. [12,22]). A border pixel pair is a pair of horizontally, vertically or diagonally directly adjacent image points, which is situated across the boundary between a changed region and an unchanged one. This implies that both pixels of each border pixel pair carry different labels. As shown in e.g. [12,22], the number of border pixel pairs occurring in a change mask is low for smoothly shaped regions, whereas the occurrence of small and wriggled regions results in a steep increase of the number of border pixel pairs. Accounting for horizontally or vertically oriented border pixel pairs and diagonally oriented ones separately, the energy $E(Q)$ can be specified as

$$E(Q) = n_B B + n_C C, \tag{11}$$

where n_B denotes the number of horizontal or vertical border pixel pairs, and n_C the number of diagonal ones. The constants B and C are the so-called potentials, which, when positive, incur an energy increase for each border pixel pair present in a change mask Q. Combining (11) with (10) leads to an expression for the a priori probability which favours the occurrence of smooth regions.

For the derivation of a practical decision rule, it is important to note that deciding between Q_u^i (i.e. $q(i) = u$) and Q_c^i (i.e. $q(i) = c$) affects only *eight* pixel pairs, as illustrated in Fig. 6. Thus, the energy $E(Q)$ can be split into a global component E_G, which assesses all border pixel pairs except those to which pixel i belongs, and a local component $E_L(q(i))$, which comprises only those border pixel pairs to which pixel i belongs. The local energy contribution depends on how many of the eight pixel pairs depicted in Fig. 6 are *border* pixel pairs. Let us denote the number of horizontal or vertical border

pixel pairs inside this neighbourhood with $v_B(q(i))$, and the number of diagonal border pixel pairs with $v_C(q(i))$. As there are four horizontally or vertically oriented pixel pairs and four diagonally oriented ones, both these numbers range between zero and four. The local energy contribution is then given by

$$E_L(q(i)) = v_B(q(i))\,B + v_C(q(i))\,C, \tag{12}$$

with $E(Q) = E_G + E_L(q(i))$. Inserting (10)–(12) into the decision rule (7) results in

$$\overline{\Delta_i^2} \underset{u}{\overset{c}{\gtrless}} t_s + 2(E_L(q(i) = c) - E_L(q(i) = u)), \tag{13}$$

where $E_L(q(i) = c)$ and $E_L(q(i) = u)$ denote the values the local energy $E_L(q(i))$ takes when $q(i) = c$ and $q(i) = u$, respectively. Writing $E_L(q(i))$ explicitly for these cases, we get

$$\overline{\Delta_i^2} \underset{u}{\overset{c}{\gtrless}} t_s + 2[(v_B(q(i) = c) - v_B(q(i) = u))\,B + (v_C(q(i) = c) - v_C(q(i) = u))\,C]. \tag{14}$$

Exploiting the fact that the labels $q(k)$ can only take binary values, the decision rule may be further simplified. If $q(i) = c$, the number $v_B(q(i) = c)$ is identical to the number $m_B^u(i)$ of pixels which border i horizontally or vertically *and* carry the opposite label u (see Fig. 6). Conversely, $v_B(q(i) = u)$ is identical to the number $m_B^c(i)$ of direct horizontal or vertical neighbours of i with label c. Similarly, $v_C(q(i) = u)$ and $v_C(q(i) = c)$ are equal to the numbers $m_C^c(i)$ and $m_C^u(i)$ of diagonal neighbours of pixel i with label c and u, respectively. Since

$$m_B^c(i) + m_B^u(i) = 4, \qquad m_C^c(i) + m_C^u(i) = 4, \tag{15}$$

decision rule (14) can be expressed as

$$\overline{\Delta_i^2} \underset{u}{\overset{c}{\gtrless}} t_s + 8(B + C) - 4(m_B^c(i)B + m_C^c(i)\,C) = \hat{t}(m_B^c(i), m_C^c(i)). \tag{16}$$

The threshold $\hat{t}(m_B^c(i), m_C^c(i))$ thus adapts to the label constellation in the neighbourhood of the considered pixel i. The higher the numbers $m_B^c(i)$ and $m_C^c(i)$ of changed neighbours, the lower is the value of the threshold, hence increasingly favouring the decision $q(i) = c$. The lowest value for the threshold is $\hat{t}(4, 4) = t_s - 8(B + C)$, and the highest one is $\hat{t}(0, 0) = t_s + 8(B + C)$. If $m_B^c(i) = m_C^c(i) = 2$, there are as many changed as unchanged neighbours, and the threshold reduces to $\hat{t}(2, 2) = t_s$. As the fixed portion t_s of the threshold thus lies in the centre of the range covered by the threshold variation, we term t_s the 'anchor threshold'.

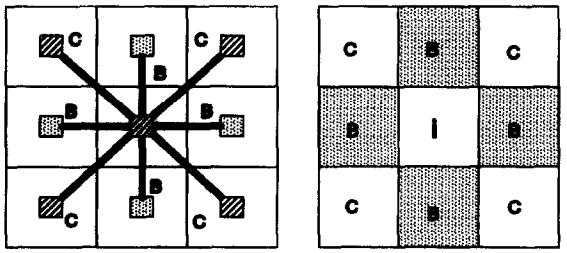

Fig. 6. Left: The eight pixel pairs examined by the local energy $E_L(q(i))$, depicted as black bars. Each pixel pair is marked with the potential B or C it incurs on $E_L(q(i))$ if it is situated across a boundary. Right: Situation depicted for decision rule (16): $m_B^c(i)$ denotes the number of shaded pixels carrying the label c (potential B), and $m_C^c(i)$ is the number of nonshaded pixels with label c (potential C).

3. Implementation: non-iterative multiple-threshold algorithms

The derivation given so far started with (1) on the presupposition that the labels $q(k)$ in the neighbourhood of the pixel i to be processed are known. In practice, this is naturally not the case. A possible way out of this dilemma would be to determine an initial change mask with a fixed, non-adaptive threshold which is then refined iteratively (cf. [5]). When working on an image sequence, however, the computational burden of iterative postprocessing can be avoided by exploiting the similarity between subsequent frames of the sequence and the corresponding similarity between subsequent change masks. As an example, we examine the computation of the change mask Q for the nth frame of an image sequence, where the image grid is scanned pixel by pixel from its upper left to its lower right corner. At this instance, the change mask $R = \{r(k)\}$ for the previous frame $n - 1$ has already been determined. When processing pixel i, the

labels $q(k)$ of its neighbours situated to the left and above have already been established (causal neighbourhood, shown shaded in Fig. 7). The labels $q(k)$ of pixels situated in the non-causal portion of the neighbourhood are not yet known. The unknown section of the label constellation is therefore approximated by labels $r(k)$ taken from the previous change mask R. For a practical implementation it is important to note that this situation, depicted in Fig. 7, emerges automatically while replacing the old labels $r(k)$ successively with new ones during the determination of Q.

Given the potentials B and C and a significance α, the threshold $\hat{t}(m_B^c(i), m_C^c(i))$ can be determined in advance for all possible combinations of $m_B^c(i)$ and $m_C^c(i)$, and stored in a look-up table. Since a potential can be regarded as a measure of interaction energy between two pixels of a border pair, which is inversely proportional to the squared distance between the pixel centres, the potentials may be related by $C = B/2$. Based on $\alpha = 5 \times 10^{-4}$, we obtain for the square sum an anchor threshold of $t_s = 55.1$ via a χ^2-distribution of $N_w = 25$ degrees of freedom. Choosing $B = 3$ and $C = 1.5$ yields the 12 additional values for the threshold $\hat{t}(m_B^c(i), m_C^c(i))$ given in Table 1. The highest threshold value is $\hat{t}(0,0) = t_s + 8(B + C) = 91.1$, and the lowest one is $\hat{t}(4,4) = t_s - 8(B + C) = 19.1$.

The decision threshold on the right-hand side of (16) can as well be used in connection with the absolute sum $\tilde{\Delta}_i$ given in (9). In this case, a χ^2-distribution with $2N_w = 50$ degrees of freedom has to be employed to establish the anchor threshold t_s. The same significance of $\alpha = 5 \times 10^{-4}$ then

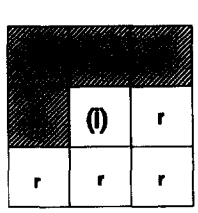

Fig. 7. Pixel i to be processed and its 3×3-neighbourhood, subdivided into a causal portion (shown shaded), and a non-causal one. When working on pixel i, the new labels $q(k)$ of its causal neighbours have already been determined, while pixels in the noncausal neighbourhood still carry 'old' labels $r(k)$ from the previous change mask R.

Table 1
Decision thresholds resulting from (16) for $\alpha = 5 \times 10^{-4}$ and $N_w = 25$. The potentials were chosen to $B = 3$ and $C = 1.5$ for the square sum $\overline{\Delta_i^2}$, and to $B = 5$ and $C = 2.5$ for the absolute sum $\tilde{\Delta}_i$ as given in (9)

$\hat{t}(m_B^c(i), m_C^c(i))$	$\overline{\Delta_i^2}$	$\tilde{\Delta}$
$\hat{t}(0,0)$	91.1	149.6
$\hat{t}(0,1)$	85.1	139.6
$\hat{t}(0,2), \hat{t}(1,0)$	79.1	129.6
$\hat{t}(0,3), \hat{t}(1,1)$	73.1	119.6
$\hat{t}(0,4), \hat{t}(1,2), \hat{t}(2,0)$	67.1	109.6
$\hat{t}(1,3), \hat{t}(2,1)$	61.1	99.6
$\hat{t}(1,4), \hat{t}(2,2), \hat{t}(3,0)$	55.1	89.6
$\hat{t}(2,3), \hat{t}(3,1)$	49.1	79.6
$\hat{t}(2,4), \hat{t}(3,2), \hat{t}(4,0)$	43.1	69.6
$\hat{t}(3,3), \hat{t}(4,1)$	37.1	59.6
$\hat{t}(3,4), \hat{t}(4,2)$	31.1	49.6
$\hat{t}(4,3)$	25.1	39.6
$\hat{t}(4,4)$	19.1	29.6

yields $t_s = 89.6$. Since the anchor threshold is in this case higher than that one obtained for the square sum, the potentials should be chosen higher as well in order to achieve the same relative threshold variation. With $B = 5$ and $C = 2.5$, the thresholds given in the right-most column of Table 1 result.

The above values for the potentials B and C were determined based on trials, with visual examination of change detection results. It also turned out that the precise parameter values are not critical, and the given threshold tables were used to process all our test sequences.

Deciding on each label $q(i)$ now reduces to the computation of the test statistic (square sum $\overline{\Delta_i^2}$ or absolute sum $\tilde{\Delta}_i$) followed by a count of changed pixels in the 3×3-neighbourhood. The appropriate decision threshold can then be retrieved from Table 1. Through the normalization of the test statistics with the noise variance σ_0^2 or the noise standard deviation σ_0, the given threshold values remain valid for sequences with different noise levels. In practice, the noise level introduced by the actual camera system in use can, for example, be estimated in advance, or recursively from unchanged regions of the sequence being processed [30, p. 202].

A further simplification of the algorithm is possible by choosing $B = C$. In this case, there is no

Table 2
Decision thresholds resulting from (17) for $\alpha = 5 \times 10^{-4}$ and $N_w = 25$. The potential was chosen to $B = 2.25$ for $\overline{\Delta_i^2}$, and to $B = 3.75$ for $\tilde{\Delta}_i$

$\hat{t}(m^c(i))$	$\overline{\Delta_i^2}$	$\tilde{\Delta}_i$
$\hat{t}(0)$	91.1	149.6
$\hat{t}(1)$	82.1	134.6
$\hat{t}(2)$	73.1	119.6
$\hat{t}(3)$	64.1	104.6
$\hat{t}(4)$	55.1	89.6
$\hat{t}(5)$	46.1	74.6
$\hat{t}(6)$	37.1	59.6
$\hat{t}(7)$	28.1	44.6
$\hat{t}(8)$	19.1	29.6

more need to discriminate between horizontal/vertical and diagonal adjacent pixels in the neighbourhood when counting changed pixels. The bivariate decision rule (16) then reduces to a univariate one:

$$\overline{\Delta_i^2} \underset{u}{\overset{c}{\gtrless}} t_s + 16B - 4B m^c(i) = \hat{t}(m^c(i)), \tag{17}$$

where $m^c(i)$ is the number of all changed pixels in the 3×3-neighbourhood of pixel i. As $m^c(i)$ varies between zero and eight, the number of different threshold values has diminished from 13 in Table 1 to only nine. In order to obtain the same minimum and maximum thresholds as in Table 1, the remaining potential B should be adjusted to the mean of the values given above, i.e. $B = 2.25$ for the square sum and $B = 3.75$ for the absolute sum. The resulting threshold values are shown in Table 2. Even if not quite in agreement with the above interaction energy interpretation, the simplified threshold array was in practice found to produce results nearly identical to those obtained from the bivariate rule (16).

For applications in block-oriented image coding it often suffices to carry out change detection blockwise instead of pixel by pixel. Our algorithm can easily be modified towards this end by simply replacing each pixel, as depicted e.g. in Figs. 6 and 7, by a block ('macro-pixel'). Doing so offers another possibility to save considerably on computation time, at the expense, however, of spatial resolution. If the spatial support of the window w_i from which the test function is computed remains unchanged, threshold Table 1 or Table 2 can still be used, otherwise, new look-up tables have to be determined.[2]

4. Results

Fig. 8 shows two change masks computed by the multiple-threshold algorithm for the speaker sequence from Fig. 1. Here, the threshold values from Table 1 were used in connection with the

[2] For blocks of size 8×8 pixels which are often used in block-oriented coding, we have $N_w = 64$. For $\alpha = 5 \times 10^{-4}$, this results in $t_s = 108$ for the square sum $\overline{\Delta_i^2}$, and $t_s = 187,3$ for the absolute sum $\tilde{\Delta}_i$.

Fig. 8. Change masks for frames no. 5 and 7 of the speaker sequence. The test statistic used was the normalized square sum $\overline{\Delta_i^2}$ in connection with the 13 thresholds from Table 1. The changed areas above the person's right shoulder are caused by moving shadow.

normalized square sum $\overline{\Delta_i^2}$. The camera noise level was estimated to $\sigma_0^2 = 4$. A comparison with the left-hand change mask from Fig. 4 makes the distinctive improvement in detection performance of the adaptive approach evident: false alarms in the background and 'holes' in the region corresponding to the moving person are both considerably reduced. A more detailed inspection reveals furthermore that the boundaries between changed and unchanged regions are now much smoother than in Fig. 4. This behaviour agrees well with the prior expectations expressed through the Gibbs/Markov model.

The same applies to the results depicted in Fig. 9, which are obtained from the traffic scene of Fig. 2. In this sequence, the noise level is much higher than in the previous one, and estimated at $\sigma_0^2 = 27$. The change masks in the first row of Fig. 9 emerged from the full range of thresholds in Table 1 being used in connection with the absolute sum $\tilde{\Delta}_i$ (9). The background now exhibits no false alarms at all, while the region detected as changed covers nearly the entire car, with only a very small number of holes present. The change masks in the second row of Fig. 9 illustrate the performance of the simplified decision rule (17): here, only the nine threshold values of Table 2 were employed. The nearly identical results confirm that in practice the simplified version of the adaptive change detector may indeed be used.

Finally, Fig. 10 shows four change masks computed from frames nos. 82, 84, 86 and 89 of the sequence *Miss America*. When comparing these results to that given in Fig. 5, the performance improvements achieved by the adaptive algorithm become strikingly evident: in contrast with Fig. 5, the background is now virtually free from detection errors, while at the same time the sporadic holes in the moving person have diminished. Test statistic and parameters were the same as those for Fig. 1. The noise level was estimated to $\sigma_0^2 = 4$.

A few words are in order here on how to treat the first frames of a sequence. In this case, no previous change mask R is available, so that we have to start with only a single threshold. Most reasonable seems using the lowest threshold value from the

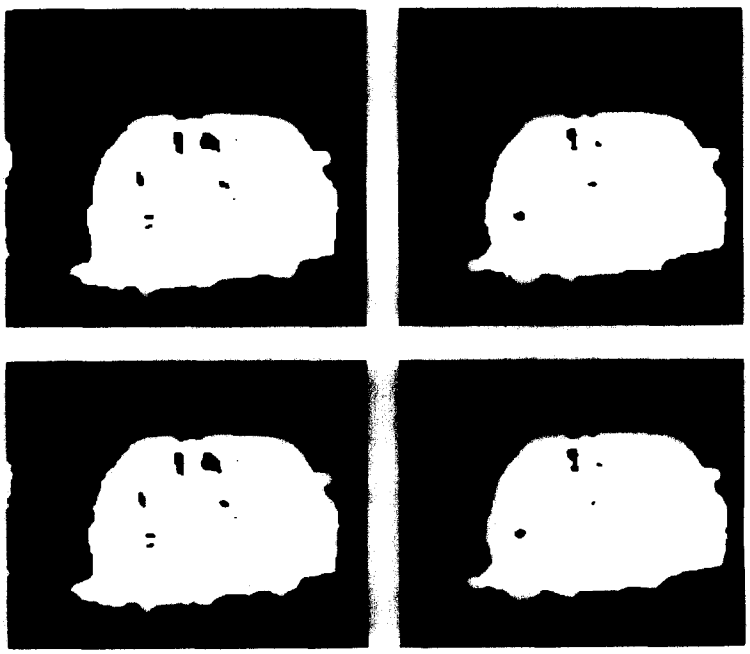

Fig. 9. First row: two change masks for the traffic scene from Fig. 2, obtained with the thresholds from Table 1. Second row: change masks for the same frames as above, but obtained using only the nine thresholds given in Table 2.

Fig. 10. Change masks for frames no. 82, 84, 86 and 89 of the sequence from Fig. 3. Test statistic and parameters as in Fig. 8.

employed look-up table (Table 1 or Table 2), or the anchor threshold t_s (i.e. $\hat{t}(2,2)$ or $\hat{t}(4)$). This choice usually results in an increased false alarm rate in the beginning, which, however, due to the erratic behaviour of these detection errors, vanishes quickly. Note that, when each frame is for instance scanned from its upper left corner to its lower right, those labels termed q in Fig. 7 are available in each neighbourhood of already the first frame, so that the algorithm starts adapting to the scene with the very first pixel 'seen'. The entire adaptation phase is usually completed after one or two frames only (In Fig. 10, for example, we started processing with frame no. 81. As can be seen, the adaptation is already nearly complete at frame no. 82). When an (undetected) scene cut has occurred, adaptation to the new image content may take a frame or two more.

5. Discussion and conclusions

The described method of change detection was developed from a Bayesian point of view, specifically from the framework of MAP estimation. There are, however, two main reasons why the change masks obtained by our algorithm are not strictly MAP estimates given the difference images D: firstly, the algorithms are of deterministic nature. Thus, they almost certainly do not find the global optimum of the posterior probability, but only a local one. Additionally, one could at least theoretically think of scanning the image grid several times for each frame until convergence is reached instead of only once. This reasoning may hold in particular for the first frames of a sequence, as otherwise generally good initializations for the optimization procedure are given by the previous change mask R. It turned out, however, that in general a single scan is sufficient to obtain a stable change mask. This is at least partly due to the fact that change detection as a *binary* segmentation problem possesses a much smaller solution space than other inverse problems of low level vision, which is why the optimization is comparatively less difficult.

The second main reason why the algorithms do not yield strict MAP-estimates is the way the

anchor threshold t_s is determined. As shown in (7), the anchor threshold t_s depends on the threshold t which must be equal to one in order to obtain MAP estimates. For practical reasons, however, t_s is determined depending on a prespecified level α. The resulting value for t_s is generally not consistent with $t = 1$. This inconsistency, however, does not affect the Bayesian reasoning underlying our considerations, as it only modifies the form of the Bayes risk subject to which the estimation is carried out. (For MAP estimation, the Bayes risk is based on a zero-one loss function [28, 24, 1].)

An earlier context-adaptive technique of change detection for image coding is 'thresholding with hysteresis' described in [10], where decision logic favours change once previous changes have occurred. Compared to that method, our approach puts adaptive change detection on a more mathematical basis. Also, for each pixel the dependence of the outcome of the decision process on the grey level differences $d(k)$ is simpler and more direct in our approach, since the method in [10] first determines a set of preliminary binary decisions, which are then combined to find the final decision. Another point where our method differs from that of [10] is that its method, unlike the one being described here, requires the computation of two test functions from the grey level difference samples, one of which emphasizes the effects of edges moving horizontally, while the other one enhances the signal-to-noise ratio of flat, moving areas.

Another Bayesian approach to change detection has been published in [27]. The resulting technique, however, consists of three processing stages (grey level difference thresholding, ML classification and MAP estimation). Of these stages, the last one is solved iteratively by the ICM algorithm, requiring several raster scans instead of only one as in our method.

A final important point is that, unlike morphological postprocessing operations such as median filtering or small region elimination, the approach described here does not force the solutions to comply with the prior knowledge at any price, but only encourages the emergence of change masks with the mentioned properties. In comparison with non-adaptive methods, our algorithm increases computational requirements only slightly, but produces greatly improved results. Considering that change masks generated by the described approach almost never need to be postprocessed, the computational costs necessary to perform our algorithm may even be less than those required for conventional non-adaptive thresholding followed by morphological postprocessing. Even when, in case of excessively noisy images, postprocessing is necessary, it will consume less computation time and produce better results for change masks obtained by the proposed algorithm due to the lower number of errors in these masks.

Acknowledgements

The original speaker sequence of Fig. 1 was recorded and kindly provided by the Research Institute of the DBP Telekom, Darmstadt, Germany. We are furthermore grateful to our former colleague R. Mester for numerous discussions, his comments on a draft of this paper and making the traffic sequence available.

The work described herein was carried out at Aachen University of Technology (RWTH). We would like to thank Prof. H.D. Lüke, director of the Institute for Communication Engineering, for his support and advice. Finally, we gratefully acknowledge the support of this work by the Robert Bosch GmbH, Hildesheim, Germany.

Appendix A

Assuming the grey level differences as being statistically independent given the alternative hypothesis H_1, as done in Section 2, is clearly a strong simplification of the situation actually encountered in changed areas, which neglects the correlations definitely present in these areas (cf. [11]). Hence, test functions like local square sum or absolute sum lead to only suboptimal evaluation of each local sample $\boldsymbol{d}_i$ [5, p. 168]. In accordance with [15], however, we argue here that significant improvements can be achieved by taking into account context rather than by trying to further optimize methods of sample evaluation. To support this view experimentally, this section examines what can be gained,

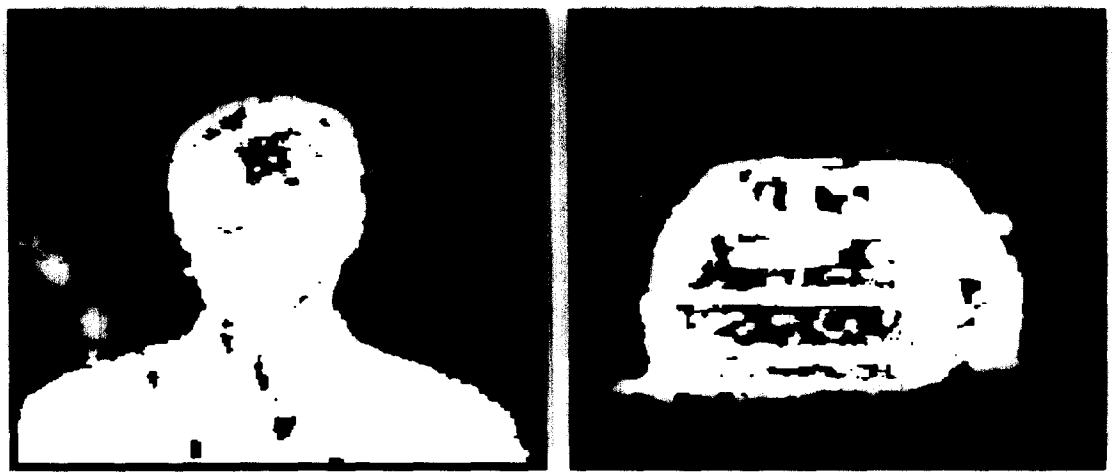

Fig. 11. Change masks obtained using the test statistic (A.2) evaluating correlations in connection with a fixed threshold value of $t = 7,5$.

if anything, by a modified test function, which estimates and evaluates the correlation function of the grey level difference process. The correlation function between two differences separated by the displacement τ is estimated by

$$\hat{\phi}_i(\tau) = \frac{1}{N_w} \sum_{k \in w_i} d(k)\, d(k+\tau), \tag{A.1}$$

with N_w being the size of the window w_i in pixels. Note that $\tau = (0, 0)$ leads to the local square sum, i.e. $\hat{\phi}_i(0, 0) = \sigma_0^2 / N_w \overline{\Delta_i^2}$. Our modified test function is formed by summing the magnitudes of estimated correlations according to

$$\bar{\phi}_i = \frac{1}{\sigma_0^2} \sum_{\tau \in N_\tau} |\hat{\phi}_i(\tau)|, \tag{A.2}$$

with the sum covering the set of displacements given by $N_\tau = \{(0,0), (0,1), (1,0), (1,1), (-1,1)\}$. The idea behind this approach is to exploit the correlations in changed areas for better separation from unchanged areas, where the differences $d(k)$ may well be assumed as independent. Change masks obtained by comparing $\bar{\phi}_i$ with a non-adaptive threshold are depicted in Fig. 11. Both these results were computed with the same threshold, with its value adjusted such that the background is nearly free of false alarms. A comparison with the results produced using the test statistic $\overline{\Delta_i^2}$ (Fig. 4) shows that there are at best minor improvements. In particular, performance in critical areas like the speaker's forehead in Fig. 4 has remained unchanged. Although no proof in the strict mathematical sense, these experiments strongly support the assumption that (carefully) neglecting certain properties of the difference data is only a minor 'offence' compared to ignoring global properties of the expected solutions. Further evidence corroborating this reasoning can be found in [1].

References

[1] T. Aach, *Bayes-Methoden zur Bildsegmentierung, Änderungsdetektion und Verschiebungsvektorschätzung*, Fortschrittberichte VDI Reihe 10, Nr. 261, VDI Verlag, Düsseldorf, 1993. Dissertation, RWTH Aachen.

[2] T. Aach and H. Dawid, "Region oriented 3D-segmentation of NMR-datasets: A statistical model-based approach", in: M. Kunt, ed., *Proc. Visual Communications and Image Processing 90*, SPIE Vol. 1360, Lausanne, October 1990, pp. 696–701.

[3] T. Aach, U. Franke and R. Mester, "Top-down image segmentation using object detection and contour relaxation", *Proc. Internat. Conf. Acoust. Speech Signal Process. 89*, Glasgow, UK, May 1989, pp. 1703–1706.

[4] T. Aach and A. Kaup, "Disparity-based segmentation of stereoscopic foreground/background image sequences", *IEEE Trans. Comm.*, Vol. 42, No. 2, 1994, pp. 673–679.

[5] T. Aach, A. Kaup and R. Mester, "Statistical model-based change detection in moving video", *Signal Processing*, Vol. 31, No. 2, March 1993, pp. 165–180.

[6] T. Aach, A. Kaup and R. Mester, "Change detection in image sequences using Gibbs random fields", *Proc. IEEE Internat. Workshop on Intelligent Signal Processing and Communication Systems*, Sendai, October 1993, pp. 56–61.

[7] M.A. Bertero, T. Poggio and V. Torre, "Ill-posed problems in early vision", *Proc. IEEE*, Vol. 76, No. 8, 1988, pp. 869–889.

[8] P. Bouthemy and P. Lalande, "Motion detection in an image sequence using Gibbs distributions", *Proc. Internat. Conf. Acoust. Speech Signal Process. 89*, Glasgow, UK, May 1989, pp. 1651–1654.

[9] C. Cafforio and F. Rocca, "Methods for measuring small displacements of television images", *IEEE Trans. Inform. Theory*, Vol. 22, No. 5, 1976, pp. 573–579.

[10] D.J. Connor, B.G. Haskell and F.W. Mounts, "A frame-to-frame picturephone coder for signals containing differential quantizing noise", *Bell System Technical J.*, Vol. 52, No. 1, 1973, pp. 35–51.

[11] D.J. Connor and J.O. Limb, "Properties of frame-difference signals generated by moving images", *IEEE Trans. Commun.*, Vol. 21, No. 10, 1974, pp. 1564–1575.

[12] H. Derin and W.S. Cole, "Segmentation of textured images using Gibbs random fields", *Comput. Vision Graph. Image Process.*, Vol. 35, 1986, pp. 72–98.

[13] N. Diehl, "Object-oriented motion estimation and segmentation in image sequences", *Signal Processing: Image Communication*, Vol. 3, No. 1, February 1991, pp. 23–56.

[14] G.W. Donohoe, D.R. Hush and N. Ahmed, "Change detection for target detection and classification in video sequences", *Proc. Internat. Conf. Acoust. Speech Signal Process. 88*, New York, NY, April 1988, pp. 1084–1087.

[15] R.M. Haralick, "Decision making in context", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 5, No. 4, 1983, pp. 417–429.

[16] B.K.P. Horn and B.G. Schunck, "Determining optical flow", *Artificial Intell.*, Vol. 17, 1981, pp. 185–203.

[17] M. Hötter and R. Thoma, "Image segmentation based on object oriented mapping parameter estimation", *Signal Processing*, Vol. 15, No. 3, October 1988, pp. 315–334.

[18] Y.Z. Hsu, H.-H. Nagel and G. Rekers, "New likelihood test methods for change detection in image sequences", *Comput. Vision Graph. Image Process.*, Vol. 26, 1984, pp. 73–106.

[19] K.P. Karmann, A.V. Brandt and R. Gerl, "Moving object segmentation based on adaptive reference images", in: L. Torres, E. Masgrau and M.A. Lagunas, eds., *Signal Processing V: Theories and Applications*, Barcelona, September 1990 (EUSIPCO 90), pp. 951–954.

[20] C. Lettera and L. Masera, "Foreground/background segmentation in videotelephony", *Signal Processing: Image Communication*, Vol. 1, No. 2, October 1989, pp. 181–189.

[21] J.O. Limb, R.F.W. Pease and K.A. Walsh, "Combining intraframe and frame-to-frame coding for television", *Bell System Technical J.*, Vol. 53, No. 6, 1974, pp. 1137–1173.

[22] R. Mester and U. Franke, "Statistical model based image segmentation using region growing, contour relaxation and classification", in: T.R. Hsing, ed., *Proc. Visual Communications and Image Processing 88*, SPIE Vol. 1001, Cambridge, USA, November 1988, pp. 616–624.

[23] L.B. Milstein and T. Lazicky, "Statistical tests for image tracking", *Comput. Vision Graph. Image Process.*, Vol. 7, 1978, pp. 413–424.

[24] A. Papoulis, *Probability & Statistics*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

[25] T. Poggio, "Early vision: From computational structure to algorithms and parallel hardware", *Comput. Vision Graph. Image Process.*, Vol. 31, 1985, pp. 139–155.

[26] T. Poggio, V. Torre and C. Koch, "Computational vision and regularization theory", *Nature*, Vol. 317, September 1985, pp. 314–319.

[27] K. Sauer and C. Jones, "Bayesian block-wise segmentation of interframe differences in video sequences", *Comput. Vision Graph. Image Process. Graphical Models and Image Processing*, Vol. 55, No. 2, 1993, pp. 129–139.

[28] C.W. Therrien, *Decision, Estimation, and Classification*, Wiley, New York, 1989.

[29] C.W. Therrien, T.F. Quatieri and D.E. Dudgeon, "Statistical model-based algorithms for image analysis", *Proc. IEEE*, Vol. 74, No. 4, 1986, pp. 532–551.

[30] R. Thoma and M. Bierling, "Motion compensating interpolation considering covered and uncovered background", *Signal Processing: Image Communication*, Vol. 1, No. 2, October 1989, pp. 191–212.

[31] H.J. Trussell and R.P. Kruger, "Comments on "nonstationary assumptions for gaussian models in images" ", *IEEE Trans. System Man Cybernet.*, Vol. 8, No. 7, 1978, pp. 579–582.