Proceedings

International Workshop
on
Intelligent Signal Processing
and
Communication Systems

October 27-29, 1993
Aoba Memorial Hall
Tohoku University, Sendai, Japan

Sponsored by
IEEE Communications Society
Signal Proc. and Comm. Electronics Committee
Asian & Pacific Committee
IEICE, Technical Group of Digital Signal Processing
Change Detection in Image Sequences Using Gibbs Random Fields: A Bayesian Approach

Til Aach¹, André Kaup¹, and Rudolf Mester²

¹Institute for Communication Engineering, Aachen University of Technology (RWTH), D-52056 Aachen, Germany
²Robert Bosch GmbH, Dept. C/FOH, D-31132 Hildesheim, Germany

Abstract

Conventional nonadaptive methods for change detection suffer from the dilemma of either causing lots of false alarms or missing considerable parts of nonstationary 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 with drastically improved performance is derived. The algorithm requires only a single raster scan per picture, thus increasing the computational load only slightly in comparison to conventional nonadaptive techniques.

1 Introduction

The detection and accurate localization of intensity changes between subsequent frames of image sequences is a crucial issue for a variety of tasks in image analysis [1, 2, 3] as well as in coding of moving video [4]. 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, the sequences often are first highpass-filtered temporally 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 in 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 (cf. [5]). Change detection by thresholding test functions computed from local samples of grey level differences thus is afflicted with the dilemma of either causing lots of false alarms or failing to detect considerable parts of moving areas [6].

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 regions¹. 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. By the exploitation of this kind of prior information about global region properties substantial improvements in detection performance can be gained. 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. [8]).

To find a formalism which allows taking into account prior knowledge about global region properties it is helpful to add change detection to the list of inverse problems of low level vision, which comprises issues like edge detection, optic flow, and surface reconstruction [9, 10]. 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 for 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 [11, 12, 13]. Starting from this framework, we develop an adaptive change detection algorithm, which, due to its noniterative nature, is very attractive also from a computational point.

¹This difficulty thus affects also change detection algorithms which do not directly evaluate grey level difference images, like [1, 7], since they suffer from the same problem of looking at local samples only.
of view. In the next section, we first formulate change detection as a Bayesian estimation problem, and specify its main components: likelihood ratio and a priori probabilities. From this formulation, decision rules and a proposal for a practical implementation will be derived.

2 The Bayesian Formulation

In the following, \( D = \{d(k)\} \) denotes 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 \( \Pr(Q|D) \) given the difference image \( D \) is maximized (MAP-estimate).

Let us for now assume that the values of the labels \( q(k) \) have already been determined 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 one produced by \( q(i) = c \) is termed \( Q_c^i \). The decision rule is then given by

\[
\frac{\Pr(Q_u^i|D)}{\Pr(Q_c^i|D)} \stackrel{\text{u}}{\geq} t, \tag{1}
\]

with \( t \) being a decision threshold. 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 to

\[
\frac{p(D|Q_u^i)}{p(D|Q_c^i)} \stackrel{\text{u}}{\geq} t \cdot \frac{\Pr(Q_u^i)}{\Pr(Q_c^i)}, \tag{2}
\]

where \( p(D|Q) \) denotes the likelihood of a change mask \( Q \) for the observed difference image \( D \). \( \Pr(Q_u^i) \) 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. \( p(D|Q) = \prod_k p(d(k)|q(k)) \). 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 [14], so that, at a first glance, they may not be assumed as independent. Nevertheless, we argue here that, quite in accordance with what is said in [8], a suboptimal evaluation of grey level samples may well be acceptable, as more is to be gained by proper handling of context instead of fine-tuning the signal model. In [15], experimental evidence is given which shows that for the purpose of change detection, these statistical dependencies can indeed be neglected without perceptible deteriorations in detection performance.

With this assumption, (2) can be simplified to

\[
\frac{p(d(i)|H_0)}{p(d(i)|H_1)} \stackrel{\text{u}}{\geq} t \cdot \frac{\Pr(Q_u^i)}{\Pr(Q_c^i)}, \tag{3}
\]

with \( p(d(i)|H_j) \) denoting the likelihoods for the hypotheses \( H_j, j = 0, 1 \), with respect to pixel \( i \).

For reliable detection performance, the decision should not be based on the grey level difference \( d(i) \) at pixel \( i \) only, but on a local sample \( \bar{d}_i \) comprising several differences \( d(k) \) (see e.g. [6, p. 168]). The sample \( \bar{d}_i \) is conveniently formed from the differences \( d(k) \) lying inside a small sliding window \( w_i \) centered at location \( i \). To incorporate the sample into the detection approach, (3) is slightly modified to

\[
\frac{p(d(i)|H_0)}{p(d(i)|H_1)} \stackrel{\text{u}}{\geq} t \cdot \frac{\Pr(Q_u^i)}{\Pr(Q_c^i)}. \tag{4}
\]

In contrast to (3), this rule decides on whether the null hypothesis can be accepted based on the entire sample \( \bar{d}_i = \{d(k) | k \in w_i\} \).

Let us now assume the grey level differences to obey zero-mean Gaussian distributions with variance \( \sigma_0^2 \) for \( H_0 \) and variance \( \sigma_1^2 \) for \( H_1 \). The above decision rule can now be rewritten to

\[
\exp \left\{ -\frac{1}{2} \left( 1 - \frac{\sigma_0^2}{\sigma_1^2} \right) \Delta^2 \right\} \stackrel{\text{u}}{\geq} t \cdot \left( \frac{\sigma_0}{\sigma_1} \right)^{N_w} \cdot \frac{\Pr(Q_u^i)}{\Pr(Q_c^i)}, \tag{5}
\]

with \( \Delta^2 \) being the normalized square sum

\[
\Delta^2_i = \frac{1}{2} \sigma_0^2 \cdot \sum_{k \in w_i} d^2(k). \tag{6}
\]

As changed areas typically exhibit differences of large magnitude, the variance \( \sigma_1^2 \) is much larger than the variance \( \sigma_0^2 \) caused by noise; estimates yield \( \sigma_1^2 > 100 \cdot \sigma_0^2 \). The fraction \( \sigma_0^2/\sigma_1^2 \) may thus be dropped. Taking the logarithm on both sides of (5) yields

\[
\Delta^2_i \stackrel{\text{u}}{\geq} 2 \ln \left[ t \left( \frac{\sigma_0}{\sigma_1} \right)^{N_w} \right] + 2 \ln \frac{\Pr(Q_u^i)}{\Pr(Q_c^i)}, \tag{7}
\]

where \( N_w \) denotes the size of the window \( w_i \) in picture elements. For the following investigations, a window of size 5 × 5 pixels, i.e. \( N_w = 25 \), was utilized throughout.

\footnote{The framework to be developed can similarly be evolved based on other model distributions, like the Laplacian one [6]. The choice between these model distributions has only minor effects on the outcome of the detection [15].}
The decision threshold on the right hand side of (7) consists of a fixed portion \( t_s \), which is independent from \( Q_u^t \) and \( Q_v^t \), and of a portion which depends on the a priori probabilities of the solutions. If \( Q_u^t \) 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, \( \Pr(Q_u^t) < \Pr(Q_v^t) \) results in a decreased threshold, hence favouring \( q(i) = c \).

Let us suppose for the moment that we have no prior knowledge with respect to the expected change masks. We thus have no information on whether \( Q_u^t \) or \( Q_v^t \) has higher a priori probability. This can be expressed mathematically through \( \Pr(Q_u^t) = \Pr(Q_v^t) \). Correspondingly, the logarithm of (7) vanishes, so that only the global decision threshold \( t_s \) remains. Instead of specifying \( t_s \) in terms of \( \sigma_0, \sigma_1 \) and \( t_s \) as indicated in (7), it is more practical to couple \( t_s \) to the rate \( \alpha \) of false alarms associated with the test. As the normalized square sum \( \frac{\Delta^2_i}{e} \) — given the null hypothesis \( H_0 \) — obeys a \( \chi^2 \)-distribution with \( N_\omega \) degrees of freedom, the threshold \( t_s \) can be determined from

\[
\Pr(\Delta^2_i > t_s | H_0) = \alpha ,
\]

once an acceptable false alarm rate \( \alpha \) has been chosen. This procedure is termed a significance test [16, 17, 6], with the false alarm rate \( \alpha \) being called the significance. As no prior knowledge is brought to bear, the resulting nonadaptive technique is associated with the plight of either missing considerable parts of moving objects or producing lots of false alarms. Figs. 1 and 2 show original sequences which are used as examples to demonstrate this behaviour in Figs. 3 and 4.

Figure 1: Frame 3 of a traffic scene.

Figure 2: Portion of size 256 x 256 pixels from frame 80 of the sequence Miss America.

Figure 3: Change mask obtained from Fig. 1 for \( \alpha = 10^{-6}, t_s = 74.5 \). While the background is nearly error-free, considerable portions of the moving objects could not be detected.

Figure 4: Change mask obtained from Fig. 2 for \( \alpha = 10^{-2}, t_s = 44.3 \). Due to the higher \( \alpha \), there occur numerous detection errors in the background.

2.1 The a priori probability

The change masks shown so far illustrate the already mentioned opposite properties of regions corresponding to moving objects and false alarms: while objects tend to manifest themselves as compact regions of preferably smooth shape, false alarms appear typically 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

\[
\Pr(Q) = \frac{1}{Z} \cdot \exp \left\{ -E(Q) \right\} ,
\]

with \( Z \) being a normalization constant. \( E(Q) \) is a so-called energy term, which assesses the state of the change masks \( Q \). Equation (9) 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 regarding so-called border pixel pairs associated with a change mask \( Q \) (see e.g. [11, 18]). 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. [11, 18], 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 adjacent border pixel pairs and diagonally adjacent ones separately, the energy \( E(Q) \) can be specified as

\[
E(Q) = n_B \cdot B + n_C \cdot C ,
\]

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 so-called potentials, which, when positive, incur an energy increase for each border pixel pair present in a change mask \( Q \). Combining (10) with (9) leads to an expression for the a priori probability which favours the occurrence of smooth regions.

It is now important to note that deciding between \( Q^u_1 \) (i.e. \( q(i) = u \)) and \( Q^c_1 \) (i.e. \( q(i) = c \)) affects only eight pixel pairs, as illustrated in Fig. 5. 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. 5 are border pixel pairs. Let us denote the number of horizontal or vertical border pixel pairs inside this neighbourhood with \( \nu_B(q(i)) \), and the number of diagonal border pixel pairs with \( \nu_C(q(i)) \). As there are four horizontally or vertically oriented pixel pairs and four diagonally oriented ones, both these numbers range between 0 and 4. The local energy contribution is then given by

\[
E_L(q(i)) = \nu_B(q(i)) \cdot B + \nu_C(q(i)) \cdot C ,
\]

with \( E(Q) = E_G + E_L(q(i)) \). Inserting (9), (10) and (11) into the decision rule (7), and writing \( E_L(q(i)) \) explicitly, results in

\[
\Delta^q_i = t_s + 2\left[ \left( \nu_B(q(i) = c) - \nu_B(q(i) = u) \right) \cdot B + \left( \nu_C(q(i) = c) - \nu_C(q(i) = u) \right) \cdot C \right].
\]

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 \( \nu_B(q(i) = c) \) is identical to the number \( m_B^c(i) \) of pixels which border \( (i) \) horizontally or vertically, and carry the opposite label \( u \) (see Fig. 5). Conversely, \( \nu_B(q(i) = u) \) is identical to the number \( m_B^u(i) \) of direct horizontal or vertical neighbours of \( (i) \) with label \( c \). Similarly, \( \nu_C(q(i) = u) \) and \( \nu_C(q(i) = c) \) are equal to the numbers \( m_C^u(i) \) and \( m_C^c(i) \) of diagonal neighbours of pixel \( (i) \) with label \( c \) and \( u \), respectively. Since

\[
m_B^c(i) + m_B^u(i) = 4, \quad m_C^c(i) + m_C^u(i) = 4 ,
\]

the decision rule (12) can be expressed as

\[
\Delta^q_i = t_s + 8(B + C) - 4(m_B^c(i) \cdot B + m_C^c(i) \cdot C).
\]

The threshold \( \Delta(m_B^c(i), m_C^c(i)) \) 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 are, the lower the value of the threshold is, hence increasingly favouring the decision \( q(i) = c \). If \( m_B^c(i) = m_C^c(i) = 2 \), there are as many changed as unchanged neighbours, and the threshold reduces to \( \Delta(2,2) = t_s \). In this context, we term \( t_s \) the 'anchor threshold'.

3 A noniterative multiple threshold algorithm

The derivation given so far started with (1) on the supposition 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, nonadaptive threshold which is then refined iteratively (cf. [6]).

When working on an image sequence, however, an elegant way out of this dilemma can be found by exploiting the similarity between subsequent frames of the sequence and the corresponding similarity between subsequent change masks. Let us examine the computation of the change mask \( Q \) for the \( n \)th 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.
processing pixel (i), the labels q(k) of its neighbours situated to the left and above are also already established (causal neighbourhood, shown shaded in Fig. 6). The labels q(k) of pixels situated in the noncausal 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. 6 — emerges automatically when replacing the old labels r(k) successively with new ones during the determination of Q. Of course, this procedure works as well with other scanning schedules.

Given the potentials B and C and a significance \( \alpha \), the threshold \( t(m_B^2(i), m_C^2(i)) \) can be determined in advance for all possible combinations of \( m_B^2(i) \) and \( m_C^2(i) \), and stored in a look-up table. Since the potentials can be regarded as a measure of interaction 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 \( \alpha = 5 \cdot 10^{-4} \), we obtain for the square sum an anchor threshold of \( t_s = 55.1 \) via a \( \chi^2 \)-distribution of \( N_w = 25 \) degrees of freedom. Choosing \( B = 3 \) and \( C = 1.5 \) yields the twelve additional values for the threshold \( t(m_B^2(i), m_C^2(i)) \) given in Table 1. The highest threshold value is \( t(0, 0) = t_s + 8(B + C) = 91.1 \), and the lowest one is \( t(4, 4) = t_s - 8(B + C) = 19.1 \).

Table 1

| \( t(m_B^2(i), m_C^2(i)) \) | \( t(0, 0) \) | 91,1 |
| \( t(0, 1) \) | 85,1 |
| \( t(0, 2), t(1, 0) \) | 79,1 |
| \( t(0, 3), t(1, 1) \) | 73,1 |
| \( t(0, 4), t(1, 2), t(2, 0) \) | 67,1 |
| \( t(1, 3), t(2, 1) \) | 61,1 |
| \( t(1, 4), t(2, 2), t(3, 0) \) | 55,1 |
| \( t(2, 3), t(3, 1) \) | 49,1 |
| \( t(2, 4), t(3, 2), t(4, 0) \) | 43,1 |
| \( t(3, 3), t(4, 1) \) | 37,1 |
| \( t(3, 4), t(4, 2) \) | 31,1 |
| \( t(4, 3) \) | 25,1 |
| \( t(4, 4) \) | 19,1 |

4 Results and Discussion

Fig. 7 shows two change masks computed for the traffic sequence of Fig. 1, using thresholds as given in Table 1. When comparing this result to that one of Fig. 3, the drastic improvement in performance achieved by the context-adaptive approach becomes strikingly evident. On the one hand, the stationary background of the scene is now virtually free from false alarms, while on the other hand the number of 'holes' in the region corresponding to the moving car has decreased as well. The results depicted in Fig. 8 exhibit a similar behaviour when compared to the nonadaptively obtained change mask given in Fig. 4. Again, the thresholds of Table 1 were employed.

For a final assessment of the described technique, it is important to note that our approach 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 drastically improved results. Considering that change masks generated by the described approach almost never need to be postprocessed, demands on the overall computational expense for producing valid change masks might even be decreased.

References


