Statistical Model-Based Change Detection in Moving Video

Til Aach and André Kaup and Rudolf Mester

in: Signal Processing. See also BIBTEX entry below.

document created on: December 20, 2006
created from file: sp93cdcoverpage.tex
cover page automatically created with CoverPage.sty
(available at your favourite CTAN mirror)

RWTHAACHEN
RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE AACHEN

Statistical model-based change detection in moving video

Til Aach, André Kaup

Institute for Communication Engineering, Aachen University of Technology (RWTH), W-5100 Aachen, Germany

Rudolf Mester

Robert Bosch GmbH, Dept. C/FOH, W-3200 Hildesheim, Germany

Abstract. A major issue with change detection in video sequences is to guarantee robust detection results in the presence of noise. In this contribution, we first compare different test statistics in this respect. The distributions of these statistics for the null hypothesis are given, so that significance tests can be carried out. An objective comparison between the different statistics can thus be based on identical false alarm rates. However, it will also be pointed out that the global thresholding methods resulting from the significance approach exhibit certain weaknesses. Their shortcomings can be overcome by the Markov random field based refining method derived in the second part of this paper. This method serves three purposes: it accurately locates boundaries between changed and unchanged areas, it brings to bear a regularizing effect on these boundaries in order to smooth them, and it eliminates small regions if the original data permits this.

Zusammenfassung. Ein wichtiger Punkt bei der Änderungsdetektion in Videosequenzen ist die Robustheit der Detektionsresultate gegenüber Rauschen. In diesem Beitrag werden zuerst verschiedene Teststatistiken in dieser Hinsicht miteinander verglichen. Die Verteilungen dieser Teststatistiken bei gegebener Nullhypothese werden angegeben, so daß Signifikanztests durchgeführt werden können. Ein objektiver Vergleich der verschiedenen Teststatistiken kann dann auf der Basis gleicher Fehldetektionsraten vorgenommen werden. Es werden aber auch einige Unzulänglichkeiten der aus dem Signifikanzansatz resultierenden globalen Schwellwertmethode aufgezeigt. Diese Schwächen können durch die im zweiten Teil beschriebene, auf Markov-Zufallsfeldern basierende Methode zur Verfeinerung von Änderungsmasken ausgeglichen werden. Diese Methode verfolgt drei Ziele: die Grenzen zwischen geänderten und ungeänderten Regionen werden genau lokalisiert, durch Regularisierung werden die Grenzen gegebenenfalls geglättet, und kleine Regionen werden, falls sie durch Fehldetektionen zustandegekommen sind, entfernt.

Résumé. Dans la détection des changements dans les séquences d'images il est de très grande importance d'assurer la fiabilité de détection en cas de séquences bruitées. A cet égard, la contribution compare tout d'abord des différentes fonctions à tester. Les fonctions de distribution de ces fonctions pour l'hypothèse nulle sont données ensuite; ainsi des tests de signification statistique peuvent être effectués. Une comparaison objective entre les différentes fonctions devient donc possible sur la base des taux de faux alarmes identiques. Quelques limitations des 'méthodes du seuil global' résultant de la méthode de signification seront traitées. Leurs inconvénients peuvent être surpassés par une méthode de traitement de la 'masque' de changement. Cette méthode se base sur les champs aléatoires Markoviens et sera expliquée dans la deuxieme partie de la contribution. Elle réalise trois objectifs: elle augmente la précision sur les frontières entre les régions changées et celles inchangées, elle lisse de façon adaptative ces frontières par régularisation et enfin elle élimine les régions de faible taille qui sont susceptibles de provenir d'erreurs de décision.

Correspondence to: Mr. Til Aach, Institute for Communication Engineering, Aachen University of Technology (RWTH), Melatenerstrasse 23, W-5100 Aachen, Germany.

1. Introduction

The detection of image areas where significant intensity changes occur between two subsequent frames of a sequence is an important issue in image coding (e.g. [1, 12, 33]) as well as in image analysis (e.g. [5, 6, 19, 18]). Region-oriented coding strategies of moving video, for example, require the sequences to be partitioned into regions corresponding to objects in the depicted scene. A change detector often forms an early step towards this goal, since it gives an initial rough estimate of moving image areas and static background [1, 12, 22]. In conditional replenishment coders, a change detector may similarly be helpful to limit replenishing to those areas which actually have changed [23].

In dynamic scene analysis, motion detection is often tied to change detection [13]. For instance, [6, 20, 35] describe methods for generating masks which correspond to moving objects of the scene to be analyzed, and for tracking these masks through the sequence. Input to these algorithms are, among others, temporal change detection results.

A peculiarity of change detection, however, is that it is not straightforward to formulate the problem precisely, because temporal intensity variations occur permanently. They can be due, for example, to motion or drifts in illumination, or to noise. On the other hand, it is by no means certain that a moving object necessarily causes perceivable temporal intensity variations, since the amount of temporal variation depends not only on the motion, but also on the spatial intensity gradient inside the moving region (cf. [17]). The objective of change detection is to distinguish temporal variations caused by noise from those within other categories.

Change detection algorithms used in image coding usually start with the grey value difference image between the two frames considered. The local sum (or mean) of absolute difference is computed inside a small measurement window which slides over the difference image. At each location,

this local sum of absolute differences is compared against a threshold. Whenever this threshold is exceeded, the center pixel of the current window location is marked as changed.

The crucial point here is the determination of optimal decision thresholds, allowing for minimal error probabilities, and thus guaranteeing results which are robust against noise. However, these thresholds are often arrived at empirically. In [4], two thresholds are used, introducing a third uncertain state in addition to the two states 'changed' and 'unchanged'. This third state is eliminated during a second run over the detection result. Threshold values are given for illustration, but appear to be adapted empirically to the respectively processed sequence. The change detector in [33] uses one threshold which is adapted to the variance estimate of the camera noise, but again the relationship is heuristical.

In this contribution, we describe a method for determining decision thresholds which are related to the false alarm rate associated with change detection. We obtain the decisions by methods of hypothesis testing, in particular, by significance tests. This is the appropriate approach since only the statistical properties for one hypothesis, namely the hypothesis that some observed temporal intensity variation is caused by noise only, are assumed to be known. The objective can now be formulated more precisely: a site on the image grid shall be marked as changed when evidence, collected from temporal grey level difference, does not support the assumption that the observed variation is arising from noise. The advantage of this approach is not only that we can specify the decision thresholds in terms of false alarm rates, but it also allows comparing objectively the performance of different test statistics by choosing their respective thresholds such that the (theoretical) false alarm rates are identical. These experimental comparisons give some insight about miss rates, which, due to the lacking characterization of changes, cannot be determined analytically. Three such statistics will be investigated. The first one is the local sum of squared, normalized grey level differences,

for which some justification can be given under the assumption of additive, white Gaussian camera noise. Similarly, it will be shown that the use of the second statistic – the local sum of absolute grey level differences – can be traced back to the assumption of Laplacian noise in the grey level difference. Finally, the local average of difference values is briefly examined.

Significance testing has been employed in image processing for the purpose of detecting small objects embedded in a (nonstationary) background process [30, 32]. There, detection is based on the residuals of a linear background process prediction with adaptively estimated predictor coefficients. In this context, the problem of change detection may be formulated as the detection of objects in noisy difference images. Since the noise is white, we can in this case proceed without the (causal) whitening transform associated with prediction.

A different approach to change detection which also uses significance tests has been described by Hsu et al. [18]. Without evaluating a difference image, in their contribution least-squares fitted biquadratic polynomial approximations to the image data are computed in test areas of two subsequent frames. The residual error for the case that a separate approximation has been carried out for each of the two test areas is compared against the residual error of one biquadratic approximation to the combined image data of both test areas. The decision between 'changed' and 'unchanged' for the considered test area then leads to a generalized likelihood ratio test. The pdf of this ratio can be established for the hypothesis that the image data of the test areas are compatible, so that a significance test can be carried out. Their approach thus relies on the assumption that the texture content inside each test area can be captured by a biquadratic polynomial, so that the residual error in unchanged areas may be assumed as being solely due to camera noise. An approach via the difference image avoids this assumption, since the difference image is free of texture content in the static areas,[1] thus rendering a texture model unnecessary. Furthermore, the necessity to perform the polynomial approximations makes the method computationally rather complex.

In Section 3, we describe a method for refining change detection results. It serves three main purposes. Firstly, it considerably improves the accuracy of the localization of boundaries between changed and unchanged areas, thus compensating for the blurring effect generally introduced by using measurement windows. Secondly, it smoothes these boundaries. Thirdly, and more or less as a by-product, it eliminates small, nonsignificant regions if the image data allow for it. In contrast to postprocessing methods like median filtering of the detection result and eliminating all small regions whose sizes are below a given threshold, the proposed method does not ignore the input grey value data.

2. Change detection using significance tests

2.1. Gaussian-distributed camera noise

We start by computing the grey level difference image $D = \{d_k\}$, with $d_k = y_1(k) - y_2(k)$, between the two considered pictures $Y_1 = \{y_1(k)\}$ and $Y_2 = \{y_2(k)\}$. The index k denotes the pixel locations on the image grid. Under the hypothesis that no change occurs at location k, the corresponding difference d_k obeys a zero mean Gaussian distribution $N(0, \sigma)$ with variance σ^2, that is,

$$p(d_k|H_0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{d_k^2}{2\sigma^2}\right\}. \tag{1}$$

Since the camera noise is uncorrelated between different frames, the variance σ^2 is equal to twice the variance of the assumed Gaussian camera noise distribution. H_0 denotes the null hypothesis, i.e.

[1] This holds when the uncorrupted background texture contents in both images are identical, as may well be assumed here. In other applications, like medical imaging or satellite imaging, this is often not the case, so that adaptive correlation cancellation techniques are applied [8, 14].

the hypothesis that there is no change at pixel k. Expression (1) reveals that $p(d_k|H_0)$ depends only on the squared ratio of the grey level difference normalized with its standard deviation, that is, on $(d_k/\sigma)^2$. It therefore seems reasonable to decide on the label for pixel k based on this squared ratio. The unknown parameter σ can be estimated off-line for the used camera system, or recursively on-line from unchanged regions while working on a sequence as described in [33, p. 202].

In order to make the detection more reliable, decisions like the one we are faced with are usually based on evaluating the set of differences d_i inside a small decision region instead of only a single pixel. The idea behind this approach is that by taking into account a higher number of samples, the overlap between the distributions of the test statistic for the cases that all samples are either 'unchanged' or 'changed' can be considerably decreased. We thus compute the local sum $\overline{\Delta_i^2}$ of $(d_k/\sigma)^2$ inside a small sliding window w_i, with i denoting the center pixel of the window. It is immediately evident that when assuming H_0 for all sites inside w_i, the joint distribution of the differences inside w_i depends only on $\overline{\Delta_i^2}$. Using the measurement window corresponds to applying a low-pass filter to the (squared) difference image, thus reducing the noise on the one hand (cf. [23]), but causing a blurring effect on the other hand. The reason for this is that it detects changes anywhere within the window, attributing the result to the center pixel regardless of precisely where the changes occur. Alternatively, one could as well assign the decision to all pixels inside the window. This 'macro pixel' approach considerably reduces the number of sites to be tested, but on the other hand it decreases the spatial resolution even further. An acceptable compromise is a window sized between 3×3 and 5×5 pixels. As already mentioned in the introduction, a main objective of Section 3 will be to compensate this blurring effect.

The local sum $\overline{\Delta_i^2}$ is proportional to the sample mean of $(d_k/\sigma)^2$ as computed inside the window. Under the assumption that no change occurs inside the window when centered at location i, the normalized differences d_k/σ each obey a zero-mean

Gaussian distribution[2] $N(0, 1)$ with variance 1. Thus, the sum $\overline{\Delta_i^2}$ obeys a χ^2-distribution with as many degrees of freedom as there are pixels inside the window. With the distribution $p(\overline{\Delta_i^2}|H_0)$ known, the decision between 'changed' and 'unchanged' can be arrived at by a significance test [27, 31]. For this purpose, we specify a significance level α and compute a corresponding threshold t_α according to

$$\alpha = \text{Prob}(\overline{\Delta_i^2} > t_\alpha | H_0). \qquad (2)$$

The statistic $\overline{\Delta_i^2}$ is now evaluated at each location i on the image grid, and whenever it exceeds t_α, the corresponding pixel is marked as changed, otherwise as unchanged. The significance level α may hence be interpreted as the type I error probability, that is, as the probability of rejecting H_0 although it is true.

We have thus specified the crucial decision thresholds in terms of error probabilities. In our experiments, we found that the input parameter α of the test is not critical: it may be varied between 10^{-6} and 10^{-2} without causing severe impacts on the detection results. In practice, the relationship between α and t_α may be stored in look-up tables.

2.1.1. When would the test be optimal?

Unfortunately, the above considerations leading to the square sum as a test statistic are not rigorous enough for a strict assessment of the 'goodness' of the decision procedure. The reason for this is that only the characteristics of the null hypothesis were taken into acount, whereas no precise assumptions about the alternative hypothesis were made. On the one hand, it is, of course, highly advantageous that the test could be designed without any such assumptions. On the other hand, so far nothing can be said about the type II error probability. To gain a little more insight, we briefly consider under

---

[2] Strictly speaking, this is true only if the variance estimate were replaced by the (unknown) true parameter. Assuming that σ^2 is estimated from a sample comprising N_s pixels, d_k/σ obeys a t-distribution with $N_s - 1$ degrees of freedom. Since N_s is usually rather large, the t-distribution may well be approximated by a Gaussian one.

which assumptions for H_1 the decision would be optimal in the sense of the Neyman–Pearson criterion. As is well known, the likelihood ratio $l(d_i) = p(d_i | H_0)/p(d_i | H_1)$ is a most powerful test statistic [27, p. 370]. The statistic $\overline{\Delta_i^2}$ would be most powerful if $l(d_i) = \psi(\overline{\Delta_i^2})$, i.e., for $l(d_i)$ depending only on $\overline{\Delta_i^2}$. This is for example the case for the following two assumptions: $p(d_k | H_1)$ is Gaussian distributed with zero mean and a variance $\sigma_c^2 \gg \sigma^2$, and $p(d_i | H_1) = \prod_{k \in w_i} p(d_k | H_1)$. In this case, both conditional densities $p(d_i | H_j)$, $j = 0, 1$, depend only on $\overline{\Delta_i^2}$, as does $l(d_i)$. Especially the second assumption is, however, not very realistic (see [9] for derivations of frame difference correlation in terms of the image correlation function), so that we cannot expect the above global thresholding method to produce minimal miss rate for a given significance.

2.2. Sum of absolute differences as test statistic

In this section, we derive how the methods of hypothesis testing can be adopted for the case that the local sum of absolute difference values is used as a test statistic. We thus consider

$$\Delta_i = \sum_{k \in w_i} \gamma |d_k|, \qquad (3)$$

where γ^{-1} is a normalization parameter adaptable to different noise levels. Tracing the considerations outlined above backward, let us search for the pdfs for H_0 and H_1 which depend only on Δ_i, so that this also holds for $l(d_i)$. We thus require the joint pdf for H_0 (as well as that one for H_1) to be of the form

$$p_N(\{d_k ; k \in w_i\} | H_0) = f_N\left(\sum_{k \in w_i} \gamma |d_k|\right). \qquad (4)$$

The subscript N denotes the size of the window w_i in pixels, which influences the structure of f_N (see Appendix A). In particular, we have for a single pixel ($N = 1$):

$$p_1(d_k | H_0) = f_1(\gamma |d_k|). \qquad (5)$$

It is shown in Appendix A that conditions (4) and (5) lead to the biexponential distribution (or

Laplace distribution), which is widely used, e.g. for the description of prediction error images. Hence, d_k obeys

$$p_1(d_k | H_0) = \frac{\gamma}{2} \exp\{-\gamma |d_k|\}. \qquad (6)$$

The variance σ^2 of this pdf is related to the parameter γ by

$$\sigma^2 = \frac{2}{\gamma^2}. \qquad (7)$$

The normalized variable $\delta_k = 2\gamma d_k$ is then distributed according to

$$p_\delta(\delta_k | H_0) = \frac{1}{4} \exp\left\{\frac{-|\delta_k|}{2}\right\}, \qquad (8)$$

and thus its absolute value obeys

$$p_{|\delta|}(|\delta_k| | H_0) = 2\varepsilon(\delta_k) p_\delta(\delta_k | H_0), \qquad (9)$$

with $\varepsilon(x)$ denoting the unit step function.

Equation (9) describes a χ^2-pdf with two degrees of freedom. In order to be able to use the tables for the χ^2-distribution, we replace the Δ_i of (3) by $\tilde{\Delta}_i = 2\Delta_i$, which is a sum of terms $|\delta_k|$, each of which is distributed according to (9). Thus, $p(\tilde{\Delta}_i | H_0)$ is a χ^2-pdf with *twice* as many degrees of freedom as there are samples inside the local window w_i. With this pdf known, the significance test can be performed as described above. Note again that, once the test statistic is chosen, the test does not depend on the assumptions for H_1.

What remains to be done is to calculate the pdf $p_K(x)$ of the camera noise, when the difference process is supposed to be described by (6). We have hence to find a pdf which, when convolved with itself, results in (6). This (de-)convolution is best performed using the Fourier transform; taking the square root of the characteristic function of (6) and subjecting the result to a Fourier transform yields

$$p_K(x) = \frac{\gamma}{\pi} K_0(|x|), \qquad (10)$$

with $K_0(x)$ denoting the zeroth-order Macdonald's function. This density bears some resemblance to

the biexponential density; it exhibits, however, a singularity for $x = 0$.

2.3. Other test statistics

Apart from squared and absolute differences, other test statistics have been used as well. Presumably the most important alternative approach is to compare the magnitude $|\overline{\Delta}_i| = |\sum_{k \in w_i} (d_k / \sigma)|$ of the local sum of differences (or the local mean) with a threshold. This approach is motivated by the fact that inside of regions which may be well approximated by bilinear functions, motion can lead to a spatially homogeneous 'offset' in the difference image. However, highly textured moving areas, which manifest themselves in the difference image as closely spaced spots of differences of differing sign, are prone to be overlooked by this statistic. This effect is confirmed in the results section, where $|\overline{\Delta}_i|$ is compared with the statistics discussed above (Fig. 11). Towards this end, a significance test has again been performed. Establishing the pdf $p(|\overline{\Delta}_i| \,|\, H_0)$ is straightforward, with $p((d_k/\sigma) \,|\, H_0)$ being a zero-mean Gaussian pdf with unit variance, $p(\overline{\Delta}_i | H_0)$ is a zero-mean Gaussian pdf with variance N, and $p(|\overline{\Delta}_i| \,|\, H_0)$ is given by $p(|\overline{\Delta}_i| \,|\, H_0) = 2\varepsilon(\overline{\Delta}_i)p(\overline{\Delta}_i | H_0)$. With this pdf known, thresholds can be computed from significance levels as described above.

3. MAP – change detection

As can be seen in Section 4, the change detection schemes analyzed above exhibit some shortcomings which were to be expected. First, there inevitably occur decision errors. Typically, these errors appear as small isolated spots inside otherwise correctly labeled regions. Another drawback is that in some critical image areas the boundaries between differently classified regions tend to be somewhat irregular. Since the change mask is assumed as being due to movements of usually compactly

shaped objects in the scene against static background, we would, however, rather expect smooth region boundaries.

To tackle these problems, many authors refine their change masks by eliminating all regions with size below a given threshold, and/or by applying a median filter to the binary change mask. This approach, however, appears questionable since it operates solely on the binary label field while completely disregarding the original image data, i.e. the difference field. These operations thus falsely eliminate areas which should definitely be marked as changed, but are being caused by small moving objects or by homogeneous objects exhibiting only small apparent motion. The small region elimination step is also disadvantageous from a computational point of view, since it requires that a connected component analysis be at least partly performed. Perhaps most important, however, is that these operations do not tackle a problem which is inherent to all algorithms employing some sort of measurement window. As already mentioned, using the window corresponds to a low-pass filtering operation, thus introducing a blurring effect which reduces the spatial resolution. This may impede the proper localization of the region boundaries, leading to the typical 'corona'-effects. The median filter, as applied to a change mask, is a purely morphological operation on a binary image [24], and while eliminating speckles, it even tries to preserve the region boundaries, inaccurate though they may be.

3.1. MAP estimation

To avoid these shortcomings, we propose to modify the above change masks subject to the MAP criterion. This approach enables us to influence the outcome of the detection by bringing to bear our expectations about change masks, while it still allows the data to play a role. Hence, we try to find the change mask $Q = \{q_k\}$ which maximizes the a posteriori density $p(Q|D)$, where D is the given difference image. The label q_k at location k can take either the value u for 'unchanged' or c for

'changed'. According to Bayes' rule, maximizing the a posteriori density is equivalent to maximizing the product $p(D|Q)p(Q) = p(D, Q)$, which is composed of the likelihood $p(D|Q)$ and the a priori density $p(Q)$. Here, we have taken advantage of the fact that the pdf for a given difference image is fixed, and may thus be ignored for the maximization process.

Assuming the difference values as conditionally independent, we can decompose the likelihood into the product

$$p(D|Q) = \prod_k p(d_k|q_k). \qquad (11)$$

For Gaussian camera noise, $p(d_k|q_k)$ is identical to the right-hand side of (1) for the case $q_k = u$. For the opposite case $q_k = c$, the difference value d_k may stem from one of several random processes (generally with nonzero mean), each of which describes the grey value differences inside some subregion of the changed image area. However, in order to find a similar expression as we have for the case $q_k = u$, we strongly simplify this model by describing the mixture of different subregions which constitute the changed area by just one zero-mean Gaussian process with variance σ_c^2. Instead of Gaussian pdfs for $p(d_k|q_k)$, Laplacian ones can be used as well ([7]); we do not expect the choice between them to challenge the following considerations (cf. for restoration [34]). The zero-mean assumption appears reasonable since, on the average, subregions with a positive mean in the difference image should occur with the same frequency as those with a negative mean. The variance σ_c^2 reflects these fluctuations of the mean values, and it is hence much larger than the variance σ^2 related to the camera noise. We thus have

$$p(d_k|q_k = c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left\{-\frac{d_k^2}{2\sigma_c^2}\right\}. \qquad (12)$$

The likelihood specified so far by (1), (11) and (12) depends on the observed difference values d_k, and thus enables the data to take the desired influence on the outcome of the estimation.

The a priori density $p(Q)$ should be specified such that it reflects our expectations on the change mask. In accordance with what was said above, we wish to find an expression for $p(Q)$ such that change masks Q consisting of smoothly shaped regions are more likely to occur than other ones. This is possible by modeling the masks, i.e. the spatial arrays of binary labels q_k, as samples from two-dimensional Gibbs/Markov random fields. A good and comprehensive treatment of the theory underlying these random fields can be found, e.g., in [11]. For a detailed treatment, we refer the reader to [3, 15].

As shown in [25, 26], for example, a probability measure assessing the smoothness of region boundaries can be found by considering pairs of horizontally, vertically or diagonally adjacent pixels which are situated across these boundaries. These border pixel pairs are termed *inhomogeneous cliques*, since the labels associated with the pair's pixels are different. In [26, 11] it is shown that the boundary is smoother when the number of inhomogeneous cliques associated with a region is lower. Hence, we assign to each horizontally or vertically oriented border pixel pair a positive cost term (called *potential*) B, and to each diagonally oriented inhomogeneous clique another positive potential C. Modeling the change mask Q as a second-order Gibbs/Markov random field, the a priori density $p(Q)$ is given by the exponential expression

$$p(Q) = \frac{1}{Z} \exp\{-E_Q\}, \qquad (13)$$

with

$$E_Q = n_B B + n_C C. \qquad (14)$$

E_Q is the energy associated with a particular change mask Q, and n_B and n_C denote the numbers of respectively horizontally/vertically and diagonally oriented border pixel pairs, which occur in Q. The so-called *partition function* Z is just a normalizing constant whose value will not be needed during the optimization step to be described. $p(Q)$ exhibits the desired properties: the smoother the regions of a particular change mask Q are shaped, the lower

are the numbers n_B and n_C of that mask, and the lower is in turn its energy E_Q. The lower the energy associated with that mask, the higher its probability to occur.

The potentials B and C shall reflect the interaction between the pixels of a clique, which should vary inversely with the distance between pixels. We model the interaction to be inversely proportional to the squared distance of the pixel centres of a clique, thus relating B and C by $C = B/2$.

3.2. Contour relaxation

In this section, we devise an optimization algorithm which iteratively modifies change masks according to the derived MAP criterion. As pointed out above, the major deficiences of these initial masks are inaccurate localization of region contours and their possibly irregular shapes. Thus, we focus our attention on the region contours, or more precisely, on the pixels located at the boundaries.

Let us scrutinize the situation of a border pixel k. As a result from the previous change detection, it carries a label q_k, and we wish to decide whether we should flip q_k or keep it as it is. Let Q_u denote the change mask with $q_k = u$, and Q_c the mask when $q_k = c$. We decide on $q_k = u$ when $p(Q_u|D) > p(Q_c|D)$, otherwise we decide $q_k = c$.

The only part of the likelihood $p(D|Q)$ affected by these considerations is the local contribution $p(d_k|q_k)$. The a priori density $p(Q)$ can similarly be split into a local term and a global one. This is due to the fact that the underlying Markov field implies that the probability of q_k conditioned on the rest of Q depends only on the label constellation in the neighbourhood in Fig. 1, where those cliques are depicted of which pixel k is a member. The energy E_Q is thus composed of a global portion E_G, which is not affected by q_k, and a local contribution $E_k(q_k)$. With $v_B(q_k)$ and $v_C(q_k)$ denoting the number of inhomogeneous cliques to which pixel k belongs when its label is q_k, $E_k(q_k)$ is given by

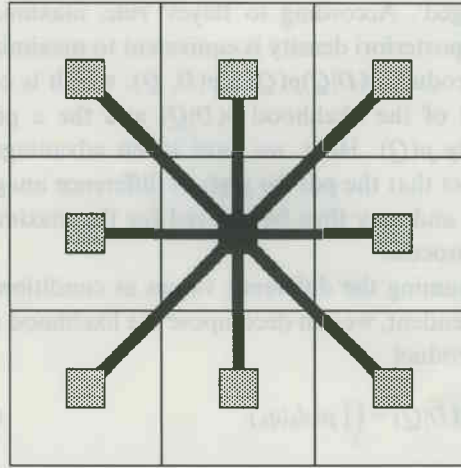$$E_k(q_k) = v_B(q_k)B + v_C(q_k)C. \tag{15}$$

Fig. 1. Neighbourhood of pixel k, with the cliques which depend on its label q_k depicted as black bars.

The decision thus reduces to

$$p(d_k|q_k = u) \exp\{-E_k(u)\}$$
$$\overset{u}{\underset{c}{\gtrless}} p(d_k|q_k = c) \exp\{-E_k(c)\}. \tag{16}$$

Exploiting (1), (12), (15) and taking the logarithm on both sides of the above inequality finally leads to the decision rule

$$d_k^2 \overset{c}{\underset{u}{\gtrless}} 2 \frac{\sigma_c^2 \sigma^2}{\sigma_c^2 - \sigma^2} \left(\ln \frac{\sigma_c}{\sigma} + (v_B(c) - v_B(u))B \right.$$
$$\left. + (v_C(c) - v_C(u))C \right). \tag{17}$$

Similarly, the Laplacian assumption (6) and (7) yields the decision rule

$$|d_k| \overset{c}{\underset{u}{\gtrless}} \frac{\sigma_c \sigma}{\sqrt{2}(\sigma_c - \sigma)} \left(\ln \frac{\sigma_c}{\sigma} + (v_B(c) - v_B(u))B \right.$$
$$\left. + (v_C(c) - v_C(u))C \right). \tag{18}$$

The right-hand side (RHS) amounts to a context dependent threshold $t(\Delta v_B, \Delta v_C)$, since it depends not only on the parameters σ_c^2, σ^2, but additionally on the differences $\Delta v_B = v_B(c) - v_B(u)$ and

$\Delta v_C = v_C(c) - v_C(u)$. The threshold thereby varies adaptively according to the labels surrounding the considered pixel k as follows. When there are more changed pixels than unchanged ones in the neighbourhood, the numbers $v_B(q_k)$, $v_C(q_k)$ of inhomogeneous cliques will be lower for the case $q_k = c$ than for $q_k = u$. This means that the differences Δv_B, Δv_C are negative, thus reducing the value of the threshold, and hence biasing the decision in favour of $q_k = c$. Similarly, the value of the threshold increases when there are more unchanged pixels in the neighbourhood, thus favouring the decision for 'unchanged'. This behaviour is in accordance with Gerhard's heuristical approach described in [16, pp. 69, 196; 21], and with the approach via 'thresholding with hysteresis' as described in [10, 23]. Furthermore, it is noteworthy that for the extreme case of very high values for parameters B and C, the decision acts similarly as a binary median filter, which assigns that label to q_k which has the majority inside the neighbourhood. For very large parameters, the RHS of (17) and (18) take negative values when more changed pixels occur in the neighbourhood, hence necessarily inducing $q_k = c$. If unchanged pixels form a majority, the threshold could become high enough to result nearly always in $q_k = u$. The decision thus reduces to a 'hard' one by majority.

Refining a given initial change mask Q is carried out by a deterministic relaxation (of the ICM type [2]) by repeatedly performing raster scans. Whenever a pixel situated at a region boundary is encountered, its label is determined according to the decision rule (17). Convergence of this local-ascent method is guaranteed, even if only to a local maximum of the a posteriori density $p(Q|D)$. This, however, is no serious drawback since the starting field provided by the previous change detection generally is good enough to ensure convergence of the contour relaxation to a reasonable result. Since the parameters σ^2 and σ_c^2 are estimated from such good starting fields, it is not necessary to update them according to the label changes.

In practice, convergence is rapid for the first few raster scans. The relaxation may hence be termin-

ated when the number of label changes per scan falls below a pre-specified level, e.g. 100 for 256×256 images. The fact that only border pixels are to be considered further speeds up the procedure.

Alternatively, the label updates may be performed in parallel as described in [2] (synchronous updating). It is of particular advantage here that the parameters σ_c^2 and σ^2 in (17) need not be updated, thus leaving only the label field Q to be affected. Since both these parameters are known before the relaxation starts, the adaptive threshold depends only on the clique differences Δv_B and Δv_C, and this relationship could be stored in a look-up table, offering another opportunity for a speed up.

4. Experimental results

In Fig. 2 the first and third frame out of a head and shoulders scene are shown, each consisting of 256×256 pixels. Figure 3 gives change masks when the local square sum inside a 5×5 window is used as test statistic. The significance level α was chosen to $\alpha = 10^{-6}$ (left), resulting in a threshold $t_\alpha = 74.5$, and to $\alpha = 10^{-2}$, resulting in a lower threshold $t_\alpha = 44.3$ (right). The variance σ^2, which is necessary for normalizing the test statistics, was estimated to $\sigma^2 = 4$. The results illustrate their robustness against variations of the input parameter α, which has been varied by several orders of magnitude without causing severe effects. In Fig. 4, we have the resulting change mask when the sum of absolute differences is employed in connection with the same significance level as in Fig. 3 (left). As could be expected, both change masks are very similar.

The following illustrations exemplify the behaviour of the contour relaxation as applied to the previous results. Figure 5 is obtained by applying the relaxation to Fig. 3. Note that the changed regions above the person's right shoulder are caused by moving shadow. The cost parameters B and C were chosen rather high in this example, to
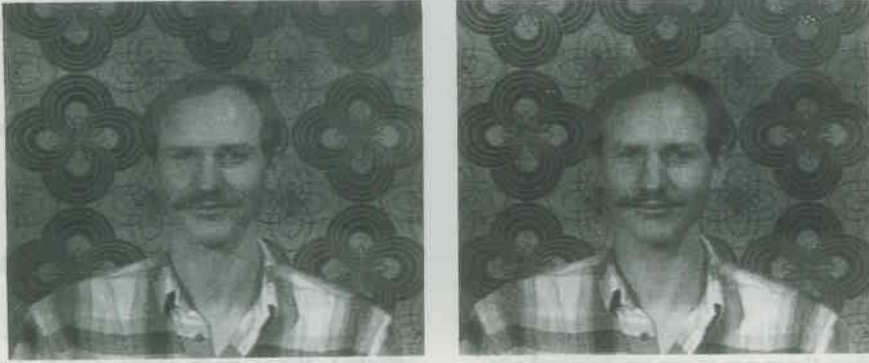
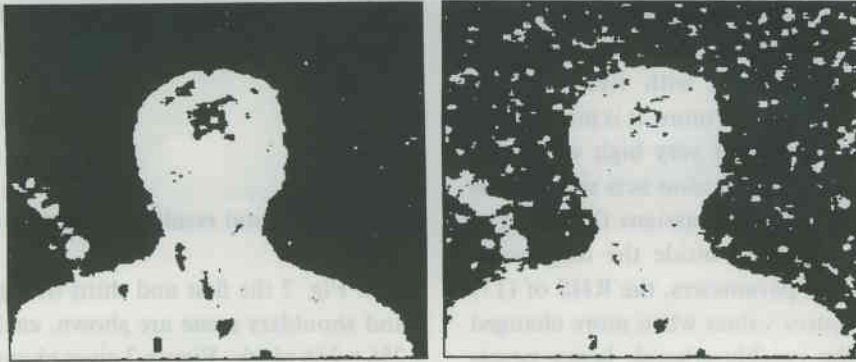Fig. 2. First and third frame of a video telephone sequence.

Fig. 3. Change masks for Fig. 2 using the local square sum. $\alpha = 10^{-6}$, $t_a = 74.5$ (left), and $\alpha = 10^{-2}$, $t_a = 44.3$ (right).

Fig. 4. Change mask obtained by employing the sum of absolute differences. $\alpha = 10^{-6}$, $t_a = 113$.

$B = 2.5$ and $C = 1.25$. In each example, the relaxation was terminated when less than 100 pixels were relabeled during a scan over the pictures. To obtain the left mask of Fig. 5 from the left one of Fig. 3,

only three scans were required, carrying out 1131 relabeling operations altogether. For the right mask, starting from the right one of Fig. 3, 5659 relabeling operations were necessary requiring six raster scans. In both cases, the variance σ_c^2 was estimated from the changed regions before the relaxation started to about $\sigma_c^2 = 400$.

A comparison between the right masks of Figs. 5 and 3 also reveals strikingly the ability of the relaxation to eliminate small, falsely detected regions. Although there is no such thing as a minimum size imposed on regions in our algorithm, most detection errors have vanished. This is in accordance with our prior expectations as expressed through the Markov model.

During our experiments, we found that the relaxation provides the best results when applied twice to an initial mask, first with relatively low
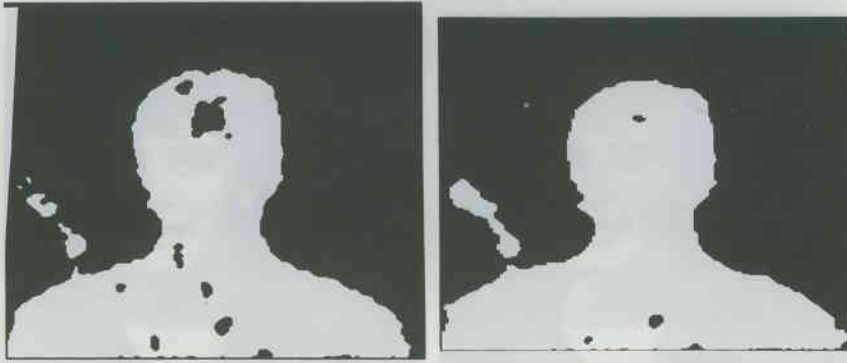
Fig. 5. Improved change masks obtained from Fig. 3 by contour relaxation. Cost parameters: $B=2.5$, $C=1.25$.

cost parameters (e.g. $B=0.5$, $C=0.25$), and subsequently with higher values $B=2.5$, $C=1.25$. This means that, during the first step, the region borders adapt particularly well to the contents of the difference image, but tend to be very irregular due to the noise. During the second run, the smoothing term dominates, thus regularizing the solution.

Figure 6 depicts a change mask which has been obtained in this way from Fig. 3 (left). The cost parameters were adjusted as just described. The first step of the relaxation resulted in 1326 relabeling operations during three scans, and the second one in 591 relabelings, again requiring three scans. To judge the accuracy of the boundary localization, Fig. 6 also shows enlarged portions of the region boundaries before and after relaxation overlaid in white over the difference image between the frames.

Figures 8 and 9 present similar results of a change detection between frames 80 and 82 of the well-known sequence *Miss America*. The variance σ^2 is nearly the same as in the previous examples. The masks in Fig. 8 are obtained with significance levels $\alpha = 10^{-6}$ and $\alpha = 10^{-2}$, using the local square sum. Applying the relaxation to the right mask in Fig. 8 yields the improved change mask in Fig. 9. Five scans were required, and a total of 4186 relabel operations were carried out. The parameter σ_c^2 was estimated to $\sigma_c^2 = 400$.

Figure 10 depicts masks obtained by the magnitude of local sum $|\overline{\Delta_i}|$ at $\alpha = 10^{-2}$ and $\alpha = 10^{-6}$. As expected, these results are not as good as the previous ones. A visual comparison with the corresponding difference image revealed that changes are particularly often overlooked along lines of
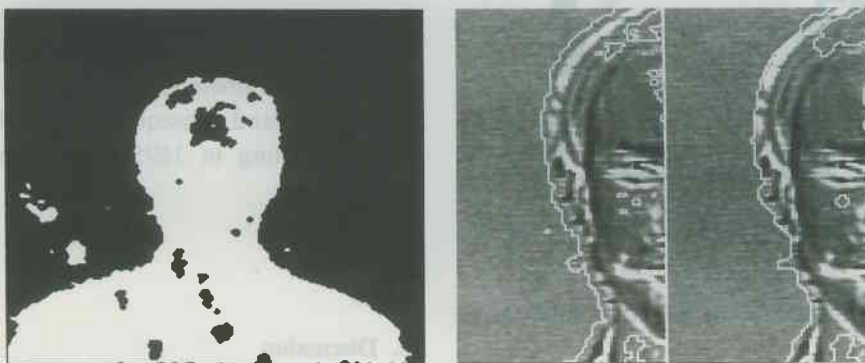

Fig. 6. Left: Improved change mask obtained from Fig. 3 (left) by a two step contour relaxation. Cost parameters: $B=0.5$, $C=0.25$ for the first step, and $B=2.5$, $C=1.25$ for the second step. Right: enlarged edge image depicting the region boundaries of this mask (right) and the raw one (left) overlaid over the difference image. The difference image has ben amplified by the factor 5.

Fig. 7. 256 × 256-pixel portions of frames 80 and 82 of the sequence *Miss America*.

Fig. 8. Change mask for Fig. 7. Test statistic: square sum, $\alpha = 10^{-6}$, $t_\alpha = 74.5$ (left), $\alpha = 10^{-2}$, $t_\alpha = 44.3$ (right).
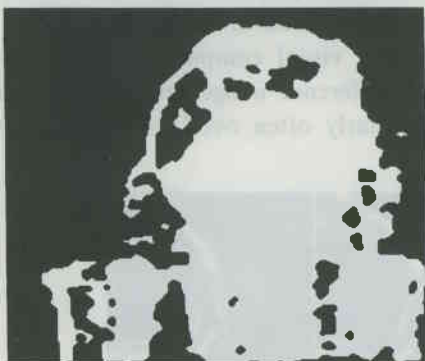

Fig. 9. Improved change mask obtained from the right one of Fig. 8 by relaxation. Cost parameters: $B = 2.5$, $C = 1.25$.

zero crossings. This behaviour is further illustrated in Fig. 11.

Finally, we present some processing examples for a traffic scene, of which Fig. 12 shows two consecutive frames. This sequence contains consider-

ably stronger noise than the previous ones; we estimated the variance to $\sigma^2 = 27$. The change masks for $\alpha = 10^{-6}$ and $\alpha = 10^{-2}$ are given in Fig. 13. Both were obtained using the sum of absolute differences as test statistic. Figure 14 shows how the relaxation improves the right mask result from Fig. 13. The relaxation was performed in two steps, first with low cost parameters, carrying out 5044 relabel operations, and subsequently with higher parameters, resulting in 1695 relabelings during four scans.

5. Discussion

The main contribution of Section 2 is a framework which allows to replace the usually empirical
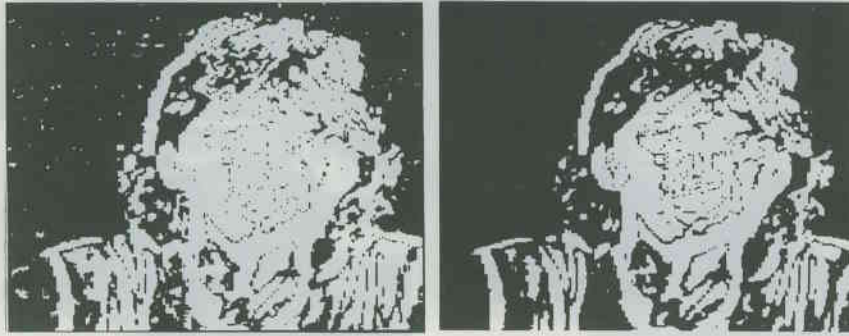
Fig. 10. Change masks obtained by employing the magnitude $|\overline{\Delta}_i|$. $\alpha = 10^{-2}$, $t_a = 12.9$ (left), $\alpha = 10^{-6}$, $t_a = 24.4$ (right).
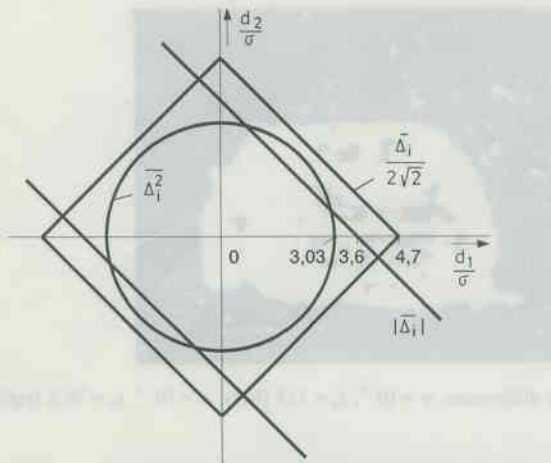

Fig. 11. Two-dimensional measurement space with decision boundaries for $\tilde{\Delta}_i^2$, $\tilde{\Delta}_i$ and $|\overline{\Delta}_i|$ when the window w_i comprises only two samples d_1/σ, d_2/σ (1×2-window). For $\alpha = 10^{-2}$, we obtain for $\tilde{\Delta}_i^2$ the circle $(d_1/\sigma)^2 + (d_2/\sigma)^2 = 9.2$ as decision boundary, the tilted square $|d_1/\sigma| + |d_2/\sigma| = 4.7$ for $\tilde{\Delta}_i$, and the parallel lines $|d_1/\sigma + d_2/\sigma| = 3.6$ for $|\overline{\Delta}_i|$. The resulting partitions of the measurement space for $\tilde{\Delta}_i^2$ and $\tilde{\Delta}_i$ are much alike, confirming similar performance of both statistics. However, the subspace for 'unchanged' of $|\overline{\Delta}_i|$ contains vectors of large magnitude which are rather unlikely to be caused by noise. On the other hand, the subspace where $|\overline{\Delta}_i|$ decides for 'changed' is nearly completely contained in the corresponding subspaces of the other two statistics. In conclusion, $|\overline{\Delta}_i|$ is hence nearly unable to detect changes which would not be detected by the other two.

threshold selection in change detectors by statistical ones which enable us to compare different statistics under controlled conditions. It turned out that the global thresholding methods could be derived by modeling the samples in the changed

areas as independent. This is, of course, only a very rough characterization. Taking the dependencies into account could possibly lead to a variable threshold, which decreases when there is evidence for correlation among the samples inside w_i, analogously as described for labels in Section 3.2. This could lead to improved performance of the detector in those areas where changes result in only weak, but nonetheless correlated difference signals.

Inherent blurring effects caused by the window are largely removed by the estimation procedure proposed in Section 3. It works on the change masks pixel by pixel, without a smearing measurement window. The approach reflects both the original image data and our a priori expectations on the detection results. In addition to improving the accuracy of the boundary localization, the proposed procedure also ensures a data-dependent small region elimination and boundary smoothing. The smoothing is due to a statistical regularization effect: the functional which is formed by the energy associated with the change masks embodies a stabilizing functional, which represents a generic constraint imposed on the desired solutions. The desired solution may thus be interpreted as one which is consistent with our prior expectations, but simultaneously is sufficiently compatible with the image data (cf. [28, p. 142; 29]). The cost parameters B and C act as regularization parameters by controlling the balance between the regularizing influence our prior expectations have on the outcome of the relaxation, and the influence of the image data.

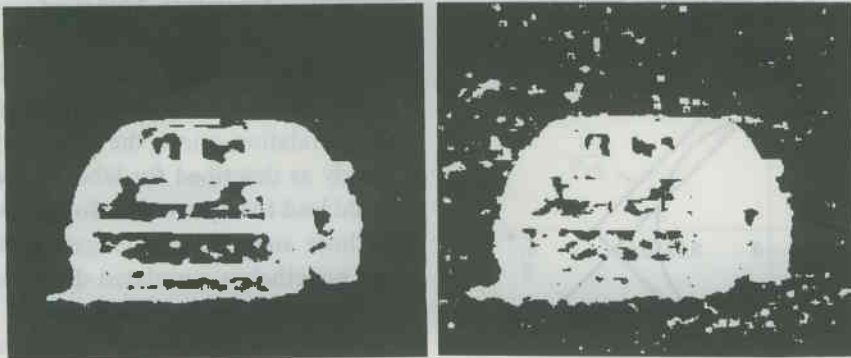Fig. 12. First and second frame of a traffic scene.

Fig. 13. Change masks for Fig. 12. Test statistc: sum of absolute differences, $\alpha = 10^{-6}$, $t_\alpha = 113$ (left), $\alpha = 10^{-2}$, $t_\alpha = 76.2$ (right).
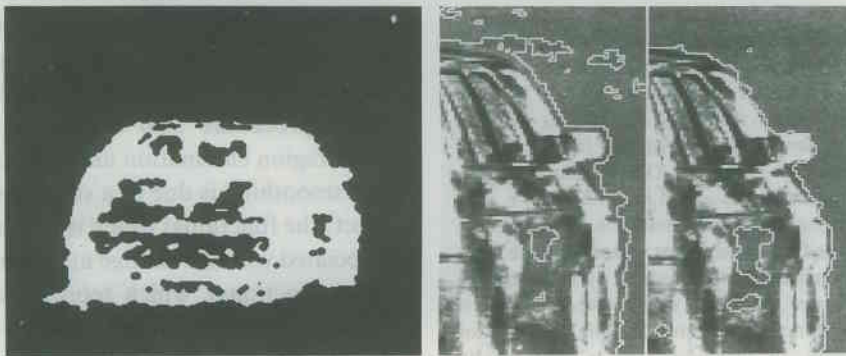

Fig. 14. Left: improved change mask obtained from the left one in Fig. 13 by a two-step relaxation. Cost parameters: $B = 0.5$, $C = 0.25$ for the first step, and $B = 2.5$, $C = 1.25$ for the second step. Right: comparison between the boundary localizations between the raw mask (left) and the improved one.

The relaxation produces its global result by way of local operations. Refining can be performed in parallel, and no connected-component analysis is required. The method is, of course, applicable to

any initial change mask, not just to one generated as described in Section 2. One could even imagine using the contour relaxation in connection with a multiresolution change detection scheme. Suppose

a change detection is carried out in a subsampled version of the grey level difference image. Depending on the degree of subsampling, the change detection speeds up substantially, but after enlarging the result to the highest resolution level, the mask will almost certainly exhibit considerable inaccuracies with respect to the boundary localization. These can be mended using the relaxation procedure, thus offering the possibility of making change detection highly efficient.

The scope of applications of the methods described in this paper is not limited to change detection. Generally, they can serve for the accurate detection of significant deviations between any two images. If one of the images is a predicted frame, generated, e.g., by motion compensation, the resulting prediction error image could also be segmented by the discussed framework.

Appendix A

Let a_k denote an observation from the camera noise in a difference image, e.g. $a_k = (d_k/\sigma)^2$ or $a_k = \gamma|d_k|$. The joint pdf for any number N of any observations shall obey

$$p_N(a_1, \ldots, a_N) = f_N\left(\sum_{i=1}^{N} a_i\right). \tag{19}$$

In particular, we have for a single observation $(N=1)$

$$p_1(a_i) = f_1(a_i). \tag{20}$$

Furthermore, we assume the noise samples as independent, yielding

$$f_N\left(\sum_{i=1}^{N} a_i\right) = \prod_{i=1}^{N} f_1(a_i). \tag{21}$$

It will be shown that conditions (19), (20) and (21) require that $p_1(a_i)$ be of the exponential form

$$p_1(a_i) = C \exp\{Aa_i\}, \tag{22}$$

with A and C denoting constants.

Taking the logarithm on both sides of (21) results in

$$\sum_{i=1}^{N} \tilde{f}_1(a_i) = \tilde{f}_N\left(\sum_{i=1}^{N} a_i\right), \tag{23}$$

where $\tilde{f} = \ln f$. Taking the partial derivative with respect to any a_j on both sides of (23) yields

$$\frac{\partial}{\partial a_j} \tilde{f}_1(a_j) = \frac{d}{d(\sum_{i=1}^{N} a_i)} \tilde{f}_N\left(\sum_{i=1}^{N} a_i\right). \tag{24}$$

It is clear that the RHS of (24) must be invariate with respect to all a_k ; $k \neq j$. This implies that $\tilde{f}_N$ is a linear function of $\sum_{i=1}^{N} a_i$, namely

$$\tilde{f}_N\left(\sum_{i=1}^{N} a_i\right) = A\left(\sum_{i=1}^{N} a_i\right) + B, \tag{25}$$

with A and B being constants. For $\tilde{f}_1(a_i)$ follows

$$\tilde{f}_1(a_i) = Aa_i + \frac{B}{N}, \tag{26}$$

yielding for $f_1(a_i)$ the exponential-type pdf

$$f_1(a_i) = \exp\{\tilde{f}_1(a_i)\} = C \exp\{Aa_i\},$$
$$C = \exp\{B/N\}. \tag{27}$$

For a Laplacian distribution, we take $a_i = \gamma|d_i|$, and $A = -1$. For a Gaussian distribution, we have $a_i = (d_i/\sigma)^2$ and $A = -\frac{1}{2}$. C follows easily in both cases, since the total probability must equal unity.

Acknowledgments

References

[1] T. Aach and A. Kaup, "Partitioning of stereoscopic sequences by evaluation of stereo disparity and temporal change detection", *Proc. Picture Coding Symposium 91*, Tokyo, Japan, September 1991, pp. 443–446.
[2] J. Besag, "On the statistical analysis of dirty pictures", *J. Royal Statist. Soc. B*, Vol. 48, No. 3, 1986, pp. 259–302.

[3] J. Besag, "Spatial interaction and the statistical analysis of lattice systems", *J. Royal Statist. Soc. B*, Vol. 36, No. 2, 1974, pp. 192–236.

[4] M. Bierling and R. Thoma, "Motion compensating field interpolation using a hierarchically structured displacement estimator", *Signal Processing*, Vol. 11, No. 4, December 1986, pp. 387–404.

[5] 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.

[6] P. Bouthemy and P. Lalande, "Detection and tracking of moving objects based on a statistical regularization method in space and time", *Proc. ECCV'90*, Antibes, France (Lecture Notes in Computer Science 427, O. Faugeras, ed., Springer, Berlin), April 1990, pp. 307–311.

[7] 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.

[8] J. Y. Chen and I. S. Reed, "A detection algorithm for optical targets in clutter", *IEEE Trans. Aerospace Electron. Systems*, Vol. 23, No. 1, 1987, pp. 46–59.

[9] D.J. Connor and J.O. Limb, "Properties of frame-difference signals generated by moving images", *IEEE Trans. Comm.*, Vol. 22, No. 10, 1974, pp. 1564–1575.

[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 Tech. J.*, Vol. 52, No. 1, 1973, pp. 35–51.

[11] 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.

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

[13] 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, 1988, pp. 1084–1087.

[14] A.S. Elfishawy, S.B. Kesler and A.S. Abutaleb, "Adaptive algorithms for change detection in image sequence", *Signal Processing*, Vol. 23, No. 2, May 1991, pp. 179–191.

[15] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No. 6, 1984, pp. 721–741.

[16] A. Gerhard, Bewegungsanalyse bei der Codierung von Bildsequenzen, Dissertation, Nachrichtentechnische Berichte, Band 19, Technische Universität München, 1988.

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

[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, ed., *Proc. EUS-IPCO 90 (Signal Processing V: Theories and Applications)*, Barcelona, Spain, September 1990, pp. 951–954.

[20] P. Lalande and P. Bouthemy, "A statistical approach to the detection and tracking of moving objects in an image sequence", in: L. Torres, E. Masgrau and M.A. Lagunas, eds., *Proc. EUSIPCO 90 (Signal Processing V: Theories and Applications)*, Barcelona, Spain, September 1990, pp. 947–950.

[21] R. Lenz and A. Gerhard, "Image sequence coding using scene analysis and spatio-temporal interpolation", in: T.S. Huang, ed., *Image Sequence Processing and Dynamic Scene Analysis*, Springer, Berlin, 1983, pp. 264–274.

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

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

[24] P. Maragos and R.W. Schafer, "Morphological filters – Part II: Their relations to median, order-statistic, and stack filters", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 35, No. 8, 1987, pp. 1170–1193.

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

[26] R. Mester, U. Franke and T. Aach, "Segmentation of image pairs and sequences by contour relaxation", in: H. Bunke, O. Kübler and P. Stucki, eds., *Proc. Mustererkennung 1988 (Informatik-Fachberichte 180)*, Springer, Zürich, September 1988, pp. 104–110.

[27] A. Papoulis, *Probability and Statistics*, Prentice Hall, Englewood Cliffs, NJ, 1990.

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

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

[30] T.F. Quatieri, "Object detection by two-dimensional linear prediction", *Proc. Internat. Conf. Acoust. Speech Signal Process. 83*, Boston, MA, 1983, pp. 108–111.

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

[32] 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.

[33] 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.

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

[35] J. Wiklund and G.H. Granlund, "Image sequence analysis for object tracking", *Proc. 5th Scandinavian Conference on Image Analysis*, Stockholm, June 1987, pp. 641–648.