A Statistical Framework for Change Detection in Image Sequences

Til Aach[1], André Kaup[1], and Rudolf Mester[2]

[1]Institute for Communication Engineering, Aachen University of Technology (RWTH), D-5100 Aachen, Germany

[2]Robert Bosch GmbH, Dept. C/FOH, D-3200 Hildesheim, Germany

Résumé

Afin d'améliorer la fiabilité dans la détection des changements dans les séquences d'images, nous décrivons deux méthodes légèrement différentes pour la détection des changements basées sur l'analyse des résultats statistiques. Cela nous permet de spécifier des seuils de décision qui sont optimales tout en respectant les probabilités d'erreur. Nous décrivons ensuite une méthode basées sur les champs Markoviens pour un traitement plus fin des résultats de la détection, méthode qui réalise trois objectifs: elle augmente la précision sur les frontières entre les régions changées et celles inchangées, elle lisse de façon adaptive ces frontières par regularisation et enfin elle élimine les régions de faible taille qui sont suceptibles de provenir d'erreurs de décision.

Abstract

To enhance the robustness of change detection, we formulate two slightly different methods for change detection as significance tests of *sufficient* statistics. This allows us to specify decision thresholds which are optimal with respect to error probability. We then describe a Markov random field based method for further processing of the detection results which serves three purposes: it increases the accuracy of the boundaries between changed and unchanged regions, it adaptively smoothes these boundaries by regularization, and it eliminates small regions in case they are likely to be caused by decision errors.

1 Introduction

The detection of image areas with significant intensity changes between two subsequent frames of a sequence is important in image coding [2, 7, 13] as well as in image analysis [6, 9]. In coding applications, change detection is used e.g. by region oriented strategies [7, 10], while in dynamic scene analysis, it serves e.g. for moving object detection and tracking.

Here, we focus on algorithms which evaluate the grey level difference image between two frames to be processed, and which are in particular wide use in coding applications. At each pixel site, the local sum (or mean) of absolute differences as computed inside a small measurement window is compared against a threshold. Whenever the threshold is exceeded, the corresponding site is marked as changed.

The crucial point here is the determination of optimal decision thresholds allowing for minimal error probabilities. However, these thresholds are often arrived at heuristically (e.g. [7, 13]).

In this contribution, we describe a method for determining decision thresholds by relating them to the false alarm rate associated with change detection. We obtain the thresholds by hypothesis testing, in particular, by significance tests. Two slightly different approaches will be analyzed: for the first one, the camera noise is assumed as additive, white and gaussian. It is shown that in this case, the local sum of ***squared normalized*** grey value differences is a sufficient statistic. For the hypothesis that no change occurs inside the local window (***null hypothesis*** H_0), the probability density function (pdf) of this test statistic can be stated, and a significance test can thus be performed. The second approach is based on regarding the already mentioned and widely used test statistic – namely, the local sum of absolute differences – as the sufficient statistic. Tracking now our previous considerations backwards, we arrive at a model for the camera noise which is implicitly assumed when using this statistic. This enables us to ascertain its pdf, and this in turn leads again to a significance test.

The described reasoning relates our approach to the proposal of [9], which also employs significance tests. However, their test statistic is quite different from ours. Without taking advantage of a difference image, the approach of [9] computes least-squares fitted biquadratic polynomials to the image data in test areas of two subsequent frames, and performs the decision by a generalized likelihood ratio test of the residual error. Their approach thus relies on the assumption that the texture content of each test area can be captured by a biquadratic polynomial, because it is only then that the residual error in unchanged areas may be assumed as being solely caused by camera noise. An approach via the difference image avoids this assumption, since the difference image is free of texture content in unchanged areas, thus rendering a texture model unnecessary. The second advantage of difference image based methods is that computationally rather complex polynomial approximations are not required.

Finally, we describe a method for refining change detection results which follows three objectives: first, it enhances the accuracy of the localization of the boundaries between changed and unchanged regions, thus compensating for blurring effects generally affiliated with the use of measurement windows. Secondly, it smoothes these boundaries. Thirdly, it eliminates small, isolated spots if they are likely to be due to decision errors. In contrast to other postprocessing methods like median filtering and/or elimination of small regions with size below a given threshold, the proposed method does not ignore the input grey value images.

2 Change Detection Using a Sufficient Statistic

2.1 Gaussian Camera Noise

We start with the grey level difference image $D = \{d_k\}$, with $d_k = y_1(k) - y_2(k)$, between two pictures $Y_1 = \{y_1(k)\}$ and $Y_2 = \{y_2(k)\}$. Under the hypothesis that no change occurs at location k, the corresponding difference d_k obeys a zero mean Gaussian distribution, i.e. $p(d_k|H_0) = N(0, \sigma)$. 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.

It is evident 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 $(d_k/\sigma)^2$. Since the camera noise is white, the joint pdf for all pixel sites inside the local measurement window w_i depends only on the sum

$$\overline{\Delta}_i = \sum_{k \in w_i} (\frac{d_k}{\sigma})^2 \qquad (1)$$

where i is the center pixel of w_i. Thus, the decision which label to assign to pixel i may be based solely on this statistic, which is therefore termed *sufficient statistic* (cf. [4]). 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 [13, p. 202].

Under the assumption that there occurs no change inside the window when centered at location i, the normalized differences d_k/σ each obey a zero mean Gaussian distribution $N(0,1)$ with variance 1. Thus, the sum $\overline{\Delta}_i$ obeys a χ^2-distribution with as many degrees of freedom as there are pixels inside the window. With the pdf $p(\overline{\Delta}_i|H_0)$ known, the decision between 'changed' and 'unchanged' can be arrived at by a significance test [4, 12]. For this purpose, we specify a significance level α, and compute the corresponding threshold t_α according to

$$\alpha = \text{Prob}\left(\overline{\Delta}_i > t_\alpha \mid H_0\right) \quad . \qquad (2)$$

Whenever $\overline{\Delta}_i$ exceeds t_α, the corresponding pixel i is marked as changed. The significance level α hence is the type I error probability, i.e. te probability to reject H_0 although it is true.

This detection method belongs to the so-called uniformly most powerful tests (one-sided in this case). Hence, once an acceptable type I error probability α has been chosen, it is guaranteed that the unknown type II error probability, i.e. the probability to miss actually changed pixels, is kept to a minimum.

Fig. 2 shows change masks obtained from a head & shoulders sequence of which Fig. 1 depicts two frames. The local square sum inside a 5 × 5-window was used in connection with $\alpha = 10^{-6}$, resulting in $t_\alpha = 74.5$ (left), and $\alpha = 10^{-2}$, resulting in $t_\alpha = 44.3$ (right).

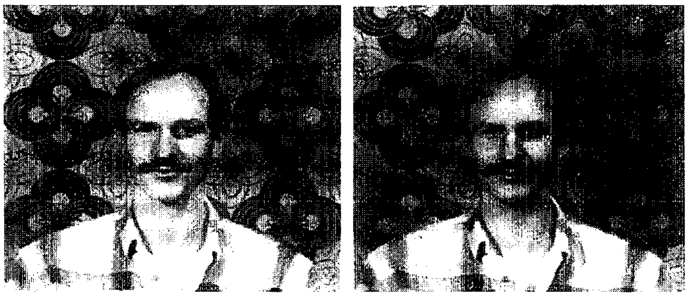

Figure 1: Original frames (256 × 256 pel, see text).

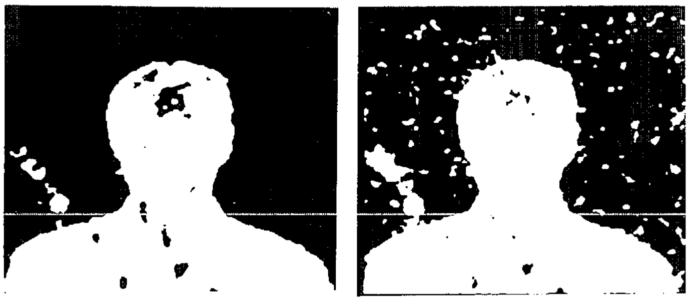

Figure 2: Change masks for Fig. 1 (see text).

2.2 Sum of absolute differences as sufficient statistic

In this section, we briefly derive a significance test assuming that the local sum of absolute difference values constitutes a sufficient test statistic. More precisely, we consider the test statistic

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

where γ^{-1} is a normalization parameter which will be specified later on. Assuming that Δ_i is sufficient, the joint pdf $p_N(\{d_k;\ k \in w_i\}|H_0)$ must depend only on Δ_i, with the subscript N denoting the size of w_i in pixels. For a single pixel ($N = 1$), it follows that $p_1(d_k|H_0)$ must depend only on $\gamma \cdot |d_i|$. As is shown in [3], these assumptions imply that $p_1(d_k|H_0)$ is Laplacian, i.e.

$$p_1\left(d_k|H_0\right) = \frac{\gamma}{2} \cdot \exp\left\{-\gamma \cdot |d_k|\right\} \quad . \qquad (4)$$

Thus, when using absolute differences, one implies that the camera noise in the difference image obeys a Laplace distribution.

It is now rather straightforward to show ([3]) that the slightly modified statistic $\tilde{\Delta}_i = 2 \cdot \Delta_i$ obeys a χ^2-distribution with *twice* as many degrees of freedom as there are pixels inside w_i. With $p(\tilde{\Delta}_i|H_0)$ known, the significance test may be performed as described.

3 Further Processing

Change detection schemes generally share the shortcoming of inevitable decision errors, which typically appear as small isolated spots inside otherwise correctly labeled regions. Furthermore, boundaries between differently classified regions often tend to be somewhat irregular. Since the change mask is assumed as being due to movements of usually compactly shaped objects, we would rather expect smooth region boundaries.

Many authors try to overcome these drawbacks by operations like median filtering the change detection result and small region elimination. Such purely morphological operations, however, are here affiliated with the disadvantage of completely ignoring the original image data. They are thus prone to errors like removing small, but correctly labelled regions. Additionally, they do not in the least tackle another serious drawback: the local window w_i introduces a blurring effect, which impedes proper localization of the region boundaries. The median filter even tries to preserve these boundaries, inaccurate though they may be.

Therefore, we propose in the following a new algorithm which doesn't exhibit these drawbacks.

3.1 MAP Estimation

To express our expectations on the change masks as well as to allow the data to play a role, we adopt the MAP approach, i.e. we try to find the change mask $Q = \{q_k\}$ which maximizes the *a posteriori* density $p(Q|D)$. The label q_k can either take the value c for 'changed' or u for 'unchanged'. Maximizing $p(Q|D)$ is equivalent to maximizing the product $p(D|Q) \cdot p(Q) = p(D,Q)$, 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 fixed probability for a given difference image may be ignored for the maximization process.

The likelihood can be decomposed into the product

$$p(D|Q) = \prod_k p(d_k|q_k) \quad . \tag{5}$$

For Gaussian camera noise, $p(d_k|q_k)$ is a Gaussian pdf for the case $q_k = u$. For $q_k = c$, the difference 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, we strongly simplify this model by describing the mixture of different subregions which form the changed area by just one zero mean Gaussian process with variance σ_c^2, i.e. $p(d_k|q_k = c)$ is given by the Gaussian pdf $N(0, \sigma_c)$. This assumption leads to a particularly elegant decision rule giving good results. The zero mean assumption is 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 hence it is much greater than the variance σ^2 related to the camera noise. Both σ^2 and σ_c^2 can be estimated from Q.

The thus specified likelihood depends on the observed difference values d_k, and hence it enables the data to influence the outcome of the postprocessing.

The a priori density $p(Q)$ shall reflect the prior knowledge of preferably smooth region boundaries. Following e.g. [11, 1], we measure smoothness by counting pairs of adjacent pixels situated across the region borders. The number of pixel pairs across a boundary is the lower, the smoother the boundary is. Making a distinction between horizontally or vertically oriented border pairs on the one hand, and diagonal ones on the other hand, we penalize each occuring border pixel pair by a positive cost term B when it is horizontal or vertical, and by another positive cost term C when it is diagonally oriented. Modelling Q as a sample from a Gibbs/Markov random field, $p(Q)$ is given by

$$p(Q) \propto \exp\{-E_Q\} \ , \quad E_Q = n_B \cdot B + n_C \cdot C \quad . \tag{6}$$

E_Q is the *energy* of a particular change mask Q, and n_B and n_C denote the numbers of respectively horizontally/vertically and diagonally oriented border pixel pairs occuring in Q. The smoother the regions of 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 of the mask, the higher its probability to occur.

The cost terms B and C shall reflect the interaction between the pixels of a clique, which should be the lower, the farther the pixels are apart. We relate them by $C = B/2$.

3.2 Contour Relaxation

Any change mask can be optimized with respect to the above MAP criterion by a deterministic relaxation, where we focus on the region contours, or more precisely, on the pixels located at the boundaries. The image grid is scanned, and for every border pixel k with its label q_k, we have to decide whether to flip q_k or leave it as it is.

Let Q_u denote the change mask when $q_k = u$, and Q_c the mask with $q_k = c$. 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 pdf $p(Q)$ can similarly be split into a local term and a global one, because 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 3×3-neighbourhood, or second order neighbourhood, of k. 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 $\nu_B(q_k)$ and $\nu_C(q_k)$ denoting the number of border pixel pairs to which pixel k belongs when its label is q_k, $E_k(q_k)$ is given by

$$E_k(q_k) = \nu_B(q_k) \cdot B + \nu_C(q_k) \cdot C \quad . \tag{7}$$

The decision thus reduces to

$$p(d_k|u) \cdot \exp\{-E_k(u)\} \ \underset{c}{\overset{u}{\gtrless}} \ p(d_k|c) \cdot \exp\{-E_k(c)\} \quad . \tag{8}$$

Exploiting (7) and taking the logarithm on both sides of the above inequality finally leads to the decision rule

$$d_k^2 \ \underset{u}{\overset{c}{\gtrless}} \ 2 \cdot \frac{\sigma_c^2 \cdot \sigma^2}{\sigma_c^2 - \sigma^2} \cdot \left(\ln \frac{\sigma_c}{\sigma} + \Delta\nu_B \cdot B + \Delta\nu_C \cdot C \right) \quad . \tag{9}$$

The right hand side amounts to a *context dependent* threshold $t(\Delta\nu_B, \Delta\nu_C)$, since it depends not only on the parameters σ_c^2, σ^2, but in addition on the differences $\Delta\nu_B = \nu_B(c) - \nu_B(u)$ and $\Delta\nu_C = \nu_C(c) - \nu_C(u)$. 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 $\nu_B(q_k)$, $\nu_C(q_k)$ of border pixel pairs will be lower for the case $q_k = c$ than for $q_k = u$. Thus, the differences $\Delta\nu_B$, $\Delta\nu_C$ are negative, what reduces the threshold, and hence favours the decision $q_k = c$. Similarly, the value of the threshold increases when there are more unchanged

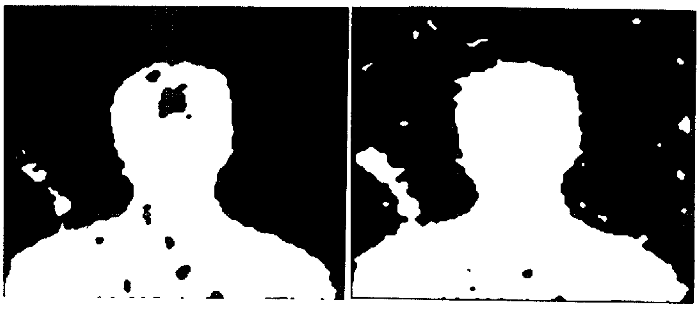

Figure 3: Relaxation results for the masks of Fig. 2 (see text).

pixels in the neighbourhood, thus favouring the decision for 'unchanged'. This behaviour agrees with the heuristic approach described in [8, p.69, p.196].
In practice, the optimization is carried out by repeated raster scans. Whenever a border pixel is encountered, its label is determined by (9). A blurring measurement window is thus not involved in this pixelwise procedure. Convergence of this method is guaranteed, even if only to a local maximum of $p(Q|D)$. This, however, is no drawback since the initial mask Q usually is good enough to ensure convergence to a reasonable result. In practice, the relaxation is terminated when the number of label changes per scan falls below a specified level, e.g. 100 for 256 × 256 images. Since only border pixels are to be considered, the procedure is also computationally advantageous. Label updating can also easily performed in parallel (synchronous updating, cf. [5]).
Fig. 3 shows the change masks of Fig. 2 modified by the relaxation, with cost parameters $B = 5$ and $C = 2.5$. A comparison to Fig. 2 strikingly reveals the ability of the relaxation to remove isolated decision errors (note that the changed areas above the person's right shoulder are caused by moving shadow).

Acknowledgement

One of the authors (T.A.) wishes to acknowledge the support granted to parts of this study by the Forschungsinstitut of the DBP TELEKOM, Darmstadt, FRG.

References

[1] T. Aach, U. Franke, R. Mester: Top-down image segmentation using object detection and contour relaxation, *Proc. ICASSP 89*, pp. 1703–1706, Glasgow, May 1989.

[2] T. Aach, A. Kaup: Partitioning of stereoscopic sequences by evaluation of stereo disparity and temporal change detection, *Picture Coding Symposium 91*, Tokyo, Sept. 1991.

[3] T. Aach, A. Kaup, R. Mester: Statistical model-based change detection in moving video, *submitted for publication.*

[4] T. W. Anderson: *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York, 1958.

[5] J. Besag: On the statistical analysis of dirty pictures, *Journal Royal Statistical Society B*, 48(3), pp. 259–302, 1986.

[6] P. Bouthemy, P. Lalande: Detection and tracking of moving objects based on a statistical regularization method in space and time, *ECCV 90*, Antibes, France, April 1990.

[7] N. Diehl: Object-oriented motion estimation and segmentation in image sequences, *Signal Processing: Image Communication*, 3, pp. 23–56, 1991.

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

[9] Y. Z. Hsu, H. Nagel, G. Rekers: New likelihood test methods for change detection in image sequences, *Computer Vision, Graphics, and Image Processing*, 26, pp. 73–106, 1984.

[10] C. Lettera, L. Masera: Foreground/background segmentation in videotelephony, *Signal Processing: Image Communication*, 1, pp. 181–189, 1989.

[11] R. Mester, U. Franke: Statistical model based image segmentation using region growing, contour relaxation and classification, *Proc. Visual Communications and Image Processing 88*, pp. 616–624, SPIE, Cambridge, USA, Nov. 1988.

[12] C. W. Therrien, T. F. Quatieri, D. E. Dudgeon: Statistical model-based algorithms for image analysis, *Proc. IEEE*, 74(4), pp. 532–551, 1986.

[13] R. Thoma, M. Bierling: Motion compensating interpolation considering covered and uncovered background, *Signal Processing: Image Communication*, 1, pp. 191–212, 1989.