Bayesian Illumination Invariant Change Detection Using a Total Least Squares Test Statistic

Til Aach[1], Rudolf Mester[2], Lutz Dümbgen[3]

[1]Institute for Signal Processing, University of Lübeck
Ratzeburger Allee 160, 23538 Lübeck, Germany

[2]Institute for Applied Physics, University of Frankfurt
Robert-Mayer-Str. 2–4, 60054 Frankfurt, Germany

[3]Institute for Mathematics, University of Lübeck
Ratzeburger Allee 160, 23538 Lübeck, Germany

aach@isip.mu-luebeck.de, mester@iap.uni-frankfurt.de, duembgen@math.mu-luebeck.de

Abstract – Changes in video data recorded by a static camera can be caused by structural scene changes like motion and by illumination changes. We describe an algorithm which discriminates reliably between structural changes and illumination, thus detecting only 'true' scene changes. To this end, we derive a new test statistic for change detection based on a Total Least Squares (TLS) approach. The basic idea is to design a test to decide whether or not two vectors observed in noise are collinear. The TLS statistic reacts to structural scene changes, while it is insensitive to varying illumination. Moreover, we integrate the TLS-statistic into a Bayesian framework for change detection, which uses *a priori* knowledge via Markov Random Fields. The resulting change detection algorithm combines the benefits of Bayesian detection with robustness against both fast and slow variations of illumination.

1 Introduction

Motion analysis for object oriented image sequence coding as well as for image interpretation frequently includes the detection and segmentation of structural scene changes in image sequences which are recorded by a static camera [1, sec. 3.3.1],[2]. Detection of the changes is often expressed in a statistical framework [3, 4, 5]. The statistical approach has the advantage that, on the one hand, the decisions can be related to detection error probabilities, usually the rate of false alarms. On the other hand, *a priori* knowledge about the expected change masks can be brought to bear to reduce the detection errors. Such Bayesian algorithms using Markov random fields (MRFs) were e.g. described in [6, 7]. These algorithms compare between the image intensities of subsequent frames, what corresponds to applying temporal highpass filter. They are therefore insensitive to slowly changing illumination. Fast changes of illumination, however, appear in the detection result, and represent detection errors from the point of view of motion detection.

Existing illumination invariant methods to change detection are often not based on statistical decision theory [8, 9], or do not use *a priori* knowledge about the sought change masks [8, 9, 10]. We therefore derive in this paper a new test statistic which is insensitive to even fast variations of illumination. Based on a model for image formation using Lambert's law, the test statistic assesses whether or not two vectors observed in noise are collinear. Moreover, we show how this approach can be integrated into the Bayesian framework of [6, 7].

2 The Image Formation Model

Roughly, the recorded image intensities can regarded as fractions of scene illumination reflected by the visible object surfaces in the direction of the camera. More precisely, each illuminated and visible object point P emits a certain power of light $L(P,d)$ per unit area (the so-called scene radiance) in the direction d of the camera. Let us develop here our image formation model from Lambert's surface model, which assumes that each surface point reflects equally well into all directions d. A surface point $P=(X,Y,Z)$ is then radiometrically described by a surface-specific scalar reflection factor $\rho(P)$, and structurally by its position P and by a unit column vector $\mathbf{n}(P)$ which is normal to the surface at P. The scene radiance $L(P)$ – which is in this case independent of d – is then given by

$$L(P) = \rho(P) \cdot I^T(P) \cdot \mathbf{n}(P) \qquad (1)$$

where $I(P)$ is a column vector representing direction and amount of the incident light, and $I^T(P) \cdot \mathbf{n}(P)$ the inner product between illumination vector and surface normal. The observed image intensities $g(m,n)$ depend on the power of light $E(m,n)$ per unit area (the so-called image irradiance) reaching pixel (m,n). According to the fundamental relationship of radiometric image formation [11], $E(m,n)$ obeys

$$E(m,n) = L(P) \cdot F^{-2} \cdot c \qquad (2)$$

where it is assumed that the 3D-point P is projected onto pixel (m,n) in the image plane. F denotes the F-number, i.e. the ratio of the camera system's focal length and its aperture [1]. Finally, c is a constant. Combining (1) and (2) yields

$$E(m,n) = \rho(P) \cdot I^T(P) \cdot \mathbf{n}(P) \cdot F^{-2} \cdot c \qquad (3)$$

Assuming that the sensor converts the image irradiance $E(m,n)$ linearly into intensities $g(m,n)$, (3) becomes

$$g(m,n) = \rho(P) \cdot I^T(P) \cdot \mathbf{n}(P) \cdot F^{-2} \cdot G \cdot c \qquad (4)$$

where G is the conversion gain. Rewriting the inner product between illumination vector and surface normal to $I^T(P) \cdot \mathbf{n}(P) = |I(P)| \cos(\alpha(P))$, where $\alpha(P)$ is the angle between surface normal and incident illumination at P, (4) becomes

$$g(m,n) = \rho(P) \cdot \cos(\alpha(P)) \cdot |I(P)| \cdot G \cdot F^{-2} \cdot c \qquad (5)$$

Assuming perspective projection, the image coordinates (m,n) depend on $P = (X,Y,Z)$ in camera centred coordinates according to [11]

$$m = \frac{f}{Z} \cdot X, \quad n = \frac{f}{Z} \cdot Y \qquad (6)$$

where f is the focal length of the camera. Since P is expressed in camera centred coordinates, Z is the distance of P to the centre of projection of the camera [2]. Inserting (6) into (5) yields

$$\begin{aligned} g(m,n) &= \rho(\frac{Z}{f}m, \frac{Z}{f}n, Z) \cdot \cos\left(\alpha(\frac{Z}{f}m, \frac{Z}{f}n, Z)\right) \\ &\quad \cdot |I(\frac{Z}{f}m, \frac{Z}{f}n, Z)| \cdot G \cdot F^{-2} \cdot c \\ &= r(m,n) \cdot a(m,n) \cdot i(m,n) \cdot G \cdot F^{-2} \cdot c \quad (7) \end{aligned}$$

with

$$\begin{aligned} r(m,n) &= \rho(\frac{Z}{f}m, \frac{Z}{f}n, Z) \\ a(m,n) &= \cos\left(\alpha(\frac{Z}{f}m, \frac{Z}{f}n, Z)\right) \\ i(m,n) &= |I(\frac{Z}{f}m, \frac{Z}{f}n, Z)| \qquad (8) \end{aligned}$$

Evidently, $r(m,n)$ depends only on the surface properties, i.e. structural scene information, whereas $i(m,n)$ depends only on the amount of illumination. The quantity $a(m,n)$ depends on the orientation of the imaged surface point relative to the incident illumination. Temporal variations in illumination at P therefore strongly affect $i(m,n)$, and not $r(m,n)$. Also, in real-world image sequences, the direction of scene illumination does usually not vary strongly over time. Then, $a(m,n)$ will also depend mainly on structural scene information. Thus, assuming temporally stable camera acquisition parameters, in particular a constant gain G and a fixed F-number F, the observed image intensities can be split into

$$g(m,n) = \hat{r}(m,n) \cdot \hat{i}(m,n) \qquad (9)$$

[1]On most lenses, the F-number is adjustable, and marked in powers of $\sqrt{2}$: $1.4, 2, 2.8, 4, 5.6, \ldots$.

[2]Eq. 6 assumes the pixel coordinates as continuous and with origin in the image centre. Discrete pixel coordinates require an additional transformation, which does not influence the following considerations.

with $\hat{r}(m,n) = r(m,n) \cdot a(m,n)$ depending almost exclusively on structural scene properties, and $\hat{i}(m,n) = i(m,n) \cdot GF^{-2}c$ depending only on illumination. Note that the multiplicative relation between these components often remains valid when the assumption of a linear conversion from image irradiance to grey levels is violated: most camera nonlinearities are decribed by a gamma-curve according to $g(m,n) = G \cdot E^\gamma(m,n)$, which still allows to split the observed grey levels according to (9).

3 The Hypothesis Test

3.1 The Test Statistic

For change detection, we compare at each pixel (m,n) the grey levels of two subsequent images which lie in a small sliding window centred at (m,n). In many cases, the illumination dependent component $\hat{i}(m,n)$ can be regarded to vary spatially only slowly [12, 13, 14, 15]. We therefore model here the illumination component $\hat{i}(m,n)$ to be spatially almost constant inside the window (cf. [9]). Thus, if no structural scene change occurs within the window, temporal differences between observed grey levels in the window are caused by a positive multiplicative factor $k(m,n)$, which accounts for possible illumination changes (see (9)), and by noise. Under this *null hypothesis* H_0, and ordering the window-internal grey values into column vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, these are related by

$$\mathbf{x}_1 = \mathbf{s} + \epsilon_\mathbf{1}, \quad \mathbf{x}_2 = k \cdot \mathbf{s} + \epsilon_\mathbf{2} \qquad (10)$$

where $\epsilon_\mathbf{i}$, $i = 1,2$, are additive noise vectors, and $\mathbf{s}$ is a signal vector. In an ideal noise free case, $\mathbf{x}_1$ and $\mathbf{x}_2$ are parallel given H_0. We thus formulate change detection as testing whether or not $\mathbf{x}_1$ and $\mathbf{x}_2$ can be regarded as collinear, with both k and $\mathbf{s}$ being unknown. Fig. 1 illus-

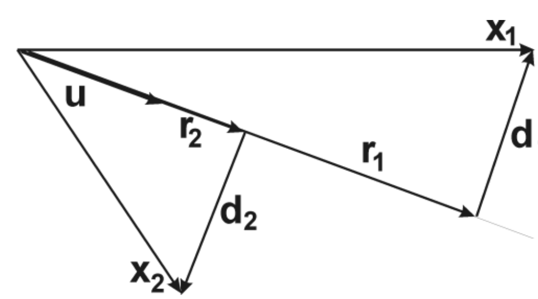


FIG. 1: Geometrical interpretation of testing the collinearity of two corrupted vectors $\mathbf{x}_1, \mathbf{x}_2$.

trates our test statistic: For the observations $\mathbf{x}_i$, $i = 1,2$ and assuming i.i.d. Gaussian noise, a maximum likelihood (ML) estimate of the true signal 'direction' (represented by unit vector $\mathbf{u}$) is obtained by minimizing the sum

$$D^2 = |\mathbf{d}_1|^2 + |\mathbf{d}_2|^2 \qquad (11)$$

of the squared distances of the observed vectors $\mathbf{x}_i$ to the axis given by vector $\mathbf{u}$. Clearly, if $\mathbf{x}_1$ and $\mathbf{x}_2$ are collinear, the difference vectors and hence the sum of their norms are zero. The projections $\mathbf{r}_i = \mathbf{x}_i^T \cdot \mathbf{u}$, $i = 1,2$ are ML estimates of the signal vectors $\mathbf{s}$ and $k \cdot \mathbf{s}$. D^2 can be rewritten to

$$D^2 = |\mathbf{x}_1|^2 + |\mathbf{x}_2|^2 - |\mathbf{x}_1^T \cdot \mathbf{u}|^2 - |\mathbf{x}_2^T \cdot \mathbf{u}|^2 \tag{12}$$

With N pixels inside the sliding window, the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ each consist of N components. With the $2 \times N$-matrix

$$\mathbf{X}^T = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \end{bmatrix} \tag{13}$$

D^2 becomes

$$D^2 = |\mathbf{x}_1|^2 + |\mathbf{x}_2|^2 - \mathbf{u}^T \cdot \mathbf{X} \cdot \mathbf{X}^T \cdot \mathbf{u} \tag{14}$$

To minimize D^2, $\mathbf{u}$ must hence be found such that $\mathbf{u}^T \cdot \mathbf{X} \cdot \mathbf{X}^T \cdot \mathbf{u}$ is maximum, subject to $\mathbf{u}^T \cdot \mathbf{u} = 1$. It is well known that this corresponds to finding the eigenvector with largest eigenvalue λ_1 of the $N \times N$-matrix $\mathbf{A} = \mathbf{X} \cdot \mathbf{X}^T$. Since we are only interested in the value of the residual sum D^2, the eigenvector $\mathbf{u}$ is not explicitly needed, as $\mathbf{u}^T \cdot \mathbf{X} \cdot \mathbf{X}^T \cdot \mathbf{u} = \lambda_1$. Hence, D^2 is then given by

$$D^2 = |\mathbf{x}_1|^2 + |\mathbf{x}_2|^2 - \lambda_1 = \mathrm{trace}(\mathbf{A}) - \lambda_1 \tag{15}$$

With the 2×2-matrix

$$\mathbf{B} = \mathbf{X}^T \cdot \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T\mathbf{x}_1 & \mathbf{x}_1^T\mathbf{x}_2 \\ \mathbf{x}_1^T\mathbf{x}_2 & \mathbf{x}_2^T\mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \tag{16}$$

we also have $\mathrm{trace}(\mathbf{A}) = \mathrm{trace}(\mathbf{B}) = \beta_1 + \beta_2$, where β_1 and β_2 are the larger and smaller eigenvalue of $\mathbf{B}$, respectively. Hence,

$$D^2 = \mathrm{trace}(\mathbf{B}) - \lambda_1 \tag{17}$$

Note now that the matrix $\mathbf{A} = \mathbf{X} \cdot \mathbf{X}^T$ is rank deficient: both $\mathbf{X}$ and $\mathbf{X}^T$ are of rank two, therefore, $\mathrm{rank}(\mathbf{A}) = 2$. Consequently, $\mathbf{A}$ has only two nonzero eigenvalues λ_1 and λ_2. Furthermore, the eigenvalues of $\mathbf{A}$ are the squared singular values of $\mathbf{X}$, which are also the eigenvalues of $\mathbf{B}$ [16, p. 55]. Therefore, $\beta_1 = \lambda_1$, and (17) becomes

$$D^2 = \beta_2 = \frac{a+c}{2} - \frac{\sqrt{(a+c)^2 - 4(ac-b^2)}}{2} \tag{18}$$

which is easily computed.

3.2 The Significance Test

Testing the null hypothesis H_0 can be stated as testing whether or not the residual sum D^2 can be explained by a given camera noise model. The difference vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ reside in a $N-1$-dimensional subspace which is orthogonal to $\mathbf{u}$. For simplicity, we model the camera noise as white, zero mean and Gaussian distributed with known variance σ_0^2. Then, if $|\mathbf{x}_1|$ and $|\mathbf{x}_2|$ are not too different, the probability density function (pdf) $p(T|H_0)$ of the test statistic $T = D^2/\sigma_0^2$ is approximately a χ^2-pdf with $N-1$ degrees of freedom. For the significance test, we specify an acceptable false alarm rate α, from which the decision threshold t is determined by

$$\alpha = \mathrm{Prob}(T > t|H_0) = \int_t^\infty p(T|H_0)dT \tag{19}$$

using a χ^2-table for $N-1$ degrees of freedom. If T exceeds t, we reject H_0, and assign the label c for changed to the window centre, otherwise, we accept H_0 and assign the label u for unchanged. This is expressed by

$$T \underset{u}{\overset{c}{\gtrless}} t \tag{20}$$

Eq. (20) leads to a global decision threshold. In the following, we show how (20) can be modified towards employing an adaptive decision threshold.

4 Adaptive MAP Change Detection

The illumination-invariant test statistic D^2 can straightforwardly be integrated into our earlier MAP-based framework for change detection [6, 7]. To this end, the sought change masks are modelled as realizations of Gibbs-Markov random fields in such a way that compactly shaped masks are preferred over ragged and noisy ones. The only additional parameter then is a so-called positive-valued potential B, which specifies interactions between the labels of adjacent pixels. The details can be found in [6, 7, 17], we give here only the algorithm.

To determine the label $q(i)$ at pixel i, i.e. to decide between $q(i) = c$ and $q(i) = u$, we first calculate the value of test statistic $D^2(i)$ from the observed grey levels inside the sliding window centred at i. We then count the number $\nu_c(i)$ of pixels labelled as "changed" in the 3×3-neighbourhood of i (see Fig. 2). Performing e.g. a

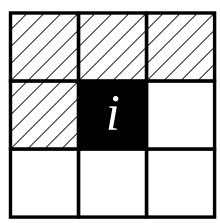


FIG. 2: 3×3-neighbourhood of a pixel i, with its causal neighbours shown shaded.

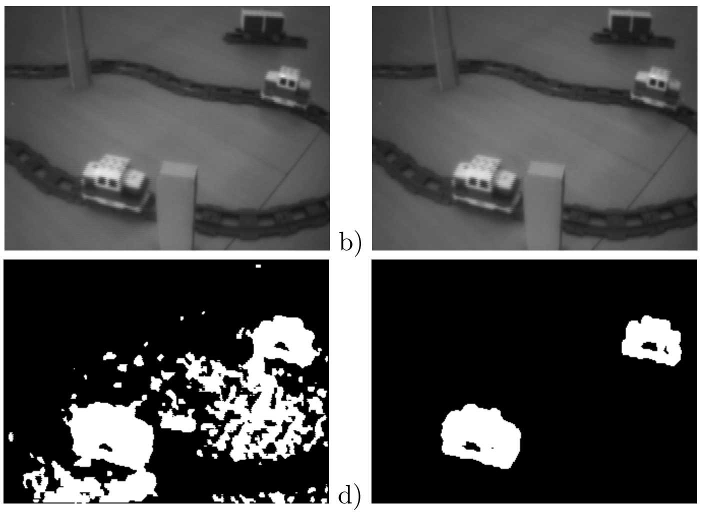
raster scan from the upper left corner to the lower right, these labels are known for the pixels left and above pixel i at the time of processing pixel i (causal neighbourhood). As estimates for the pixels in the noncausal part of the neighbourhood we simply take the labels from the previous change mask, which, in real-time image sequences, tends to be similar to the new one. Note that this constellation emerges automatically when overwriting the previous change mask pixelwise during the computation of the new one. Thus, $0 \leq \nu_c(i) \leq 8$. The adaptive decision rule is then given by

$$D^2(i) \underset{u}{\overset{c}{\gtrless}} \sigma_0^2 \cdot (t + (4 - \nu_c(i)) \cdot B) \tag{21}$$

where t is determined according to (19). The nine different values of the adaptive threshold can be precomputed and stored in a lookup table. The adaptive threshold is the lower, the larger $\nu_c(i)$, i.e. a decision for $q(i) = c$ is favoured the more, the more changed pixels are found in the immediate neighbourhood of i. Clearly, this behaviour favours the outcome of smoothly shaped change masks.

5 Results

Fig. 3 shows two successive frames of a sequence in which a beam of light is moved quickly across the scene. With a sliding window of 5×5 pixels to calculate the test statistic (i.e. $N = 25$), the illumination-sensitive algorithm of [7] evidently detects both moving objects and illumination changes. Fig. 3 d) has been computed with the adaptive version of our illumination-insensitive algorithm according to (21), which clearly detects the moving objects without reacting to the variations of illumination. Here, the significance level α was set to $\alpha = 10^{-6}$ yielding a threshold t of $t = 72.2$. The potential B was set to $B = 3.5$.



a) b) c) d)

FIG. 3: a), b): Subsequent original frames from a sequence with moving toy engines. A beam of light crosses this scene quickly from left to right. c) Result of the illumination sensitive change detection algorithm in [7], mixing illumination changes with the moving engines. d) Result of the TLS-based illumination invariant change detection. The engines were safely detected, but not the illumination changes.

6 Conclusions

Based on a tractable image formation model, we have derived a new test statistic for the detection of changes which is insensitive to variations of illumination. We also determined the distribution of this statistic given the null hypothesis, and designed a hypothesis test. We then integrated this test into an earlier adaptive framework which allows to exploit prior knowledge about typical motion masks. Future work is geared toward extensively evaluating the algorithm, and comparing it to other approaches, e.g. [17], or [14, 15], where a homomorphic prefilter was used.

References

[1] G. Tziritas, C. Labit: *Motion Analysis for Image Sequence Coding.* Elsevier, 1994.

[2] A. Mitiche, P. Bouthemy: Computation and analysis of image motion: A synopsis of current problems and methods. *Intl. Jour. Comp. Vis.* 19(1):29–55, 1996.

[3] Y. Z. Hsu, H.-H. Nagel, G. Rekers: New likelihood test methods for change detection in image sequences. *Comp. Vis. Grap. Im. Proc* 26:73–106, 1984.

[4] T. Aach, A. Kaup, R. Mester: A statistical framework for change detection in image sequences. *13ième Colloque GRETSI* 1149–1152, Juan-Les-Pins, France, Sept. 1991.

[5] T. Aach, A. Kaup, R. Mester: Statistical model-based change detection in moving video. *Sig. Proc.* 31(2):165–180, 1993.

[6] T. Aach, A. Kaup, R. Mester: Change detection in image sequences using Gibbs random fields. *Proc. IEEE Intl. Works. Intell. Sig. Proc. Comm. Sys. (ISPACS)*:56–61, Sendai, Japan, Oct. 1993.

[7] T. Aach, A. Kaup: Bayesian algorithms for change detection in image sequences using Markov random fields. *Sig. Proc.: Im. Comm* 7(2):147–160, 1995.

[8] K. Skifstad, R. Jain: Illumination independent change detection for real world image sequences. *Comp. Vis. Grap. Im. Proc.* 46:387–399, 1989.

[9] E. Durucan, T. Ebrahimi: Robust and illumination invariant change detection based on linear dependence for surveillance application. *EUSIPCO 2000*, 1041–1044, Tampere, Finland, Sept. 3-8 2000.

[10] S.-Z. Liu, C.-W. Fu, S. Chang: Statistical change detection with moments under time-varying illumination. *IEEE T IP* 7(9):1258–1268, 1998.

[11] E. Trucco, A. Verri: *Introductory Techniques for 3-D Computer Vision.* Prentice-Hall, New Jersey, 1998.

[12] A. V. Oppenheim, R. W. Schafer, T. G. Stockham Jr: Nonlinear filtering of multiplied and convolved signals. *Proc. IEEE* 56(8):1264–1291, 1968.

[13] C. R. Moloney: Methods for illumination-independent processing of digital images. *IEEE Pacific Rim Conf. Comm. Comp. Sig. Proc.*:811–814, May 9–10 1991.

[14] D. Toth, T. Aach, V. Metzler: Illumination-invariant change detection. *4th IEEE Southwest Symp. Im. Anal. Interp. (SSIAI)*:3–7, Austin, April 2 - 4 2000.

[15] D. Toth, T. Aach, V. Metzler: Bayesian spatio-temporal motion detection under varying illumination. *Eur. Sig. Proc. Conf. (EUSIPCO)*:2081–2084, Tampere, Sept. 3-8 2000.

[16] C. W. Therrien: *Discrete Random Signals and Statistical Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ, 1992.

[17] T. Aach, L. Dümbgen, R. Mester, D. Toth: Bayesian illumination-invariant motion detection. *IEEE Intl. Conf. Im. Proc. (ICIP)*:in press, Thessaloniki, Oct. 7–10 2001.