Illumination–Invariant Change Detection

Daniel Toth, Til Aach and Volker Metzler
Institute for Signal Processing, University of Lübeck
Ratzeburger Allee 160, D–23538 Lübeck, Germany
(toth, aach, metzler)@isip.mu-luebeck.de

Abstract

Moving objects in image sequences acquired by a static camera can be detected by analyzing the grey–level difference between successive frames. Direct motion detection, however, will also detect fast variations of scene illumination. This paper describes a method for motion detection that is considerably less sensitive to time–varying illumination. It is based on combining a motion detection algorithm with a homomorphic filter which effectively suppresses variable scene illumination. To this end, the acquired image sequence is modelled as being generated by an illumination and a reflectance component that are approximately separated by the filter. Detection of changes in the reflectance component is directly related to scene changes, i.e. object motion. Real video data are used to illustrate the system's performance.

1. Introduction

The detection of moving objects in image sequences recorded with a static camera is often based in evaluating temporal changes in intensity [2, 3, 4, 13]. Assuming that such temporal changes are caused by motion or noise, the purpose of change detection is to identify and label those changes which are due to motion. Ideally, the resulting change mask corresponds to the projection of moving objects and shadow onto the image plane, plus uncovered background. Two error classes occur: on the one hand, noise causes false positives in static regions (class 1 errors). On the other hand, object motion does not always generate distinct grey level changes, leading to "holes" in the change mask (class 2 errors). This is typically the case when object motion is small or spatial grey level gradients are small (cf.[12]). Below we will review a context-adaptive change detection algorithm which almost completely avoids such errors.

Temporal changes are often detected by comparing consecutive frames of the sequence in question [7, 9], for instance by calculating the grey level difference image. This corresponds to a temporal highpass filter, which more or less eliminates slow changes in illumination. Potential faster changes in illumination, however, are not sufficiently attenuated by this operation.

To eliminate the unwanted influence of a varying illumination component we use a model well known in homomorphic image filtering [11]. In this model the image intensity is considered to be generated by an incoming illumination, which is reflected by the surfaces of the objects in the observed scene. For Lambertian surfaces, the relation between observed intensity y, illumination i and reflectance r is multiplicative. Assuming the scene illumination as spatially varying slowly and the reflectance component as containing mainly spatially medium and high–frequency details [10, 11], an approximate elimination of illumination is possible by taking the logarithm of the image before applying a linear high–pass filter. After exponentiation, change detection is carried out on the remaining reflectance component.

In the following we first summarize our method for context-adaptive change detection; a more detailed description can be found in [1, 3]. This is followed by a description of our illumination–invariant change detection system.

2. The change detection algorithm

The goal of a change detection system is to generate a change mask q consisting of binary labels $q(k)$ for each pixel k on the image grid. The labels either take the value "u" ('unchanged') or "c" ('changed'). In order to determine the label $q(k = i)$ for pixel i we start with the grey–level difference image $d(k)$ between two successive frames, and compare the sum of absolute differences Δ_i within a sliding window w_i with N pixels and center i to a threshold T:

$$\Delta_i = \frac{2\sqrt{2}}{\sigma_{\mathsf{u}}} \sum_{k \in w_i} |d(k)| \underset{\mathsf{u}}{\overset{\mathsf{c}}{\gtrless}} T. \qquad (1)$$

Here, σ_u is the noise standard deviation of the grey level differences in stationary areas, which is assumed to be constant over space. Normalization by σ_u — which is known for a given camera or easily estimated — makes T insensitive to different noise levels. Given the null hypothesis H_0 (grey–level differences $d(k)$ in w_i only due to noise), and modelling the grey level differences $d(k)$ as independent and Laplacian distributed [5], Δ_i obeys a χ^2 distribution with $2N$ degrees of freedom [4] [1]. Change detection can then be formulated as a significance test, where the threshold value is determined in terms of an acceptable false alarm rate α [4]. The "significance" α is equivalent to the probability that Δ_i exceeds the threshold T, given H_0:

$$\alpha = \text{Prob}\left(\Delta_i > T | H_0\right). \tag{2}$$

For a given false alarm rate α, the threshold T is determined from tables of the χ^2-distribution. The decision rule then is

$$\Delta_i \overset{\text{c}}{\underset{\text{u}}{\gtrless}} T \quad . \tag{3}$$

Whenever Δ_i exceeds T, we decide $q(i) = c$, otherwise $q(i) = u$.

As mentioned above, this global-threshold-based decision procedure is prone to two kinds of errors, false postives and false negatives. The basic idea to reduce these is to decrease the decision threshold inside changed areas, and to increase it outside. This exploits the prior knowledge that changed and unchanged regions correspond to objects or background, which usually are of compact shape. Formally, this prior knowledge can be expressed by modelling the change masks as realizations of Gibbs/Markov-random fields [1]. This leads to a variable threshold t which adapts to the label constellation within a pixel's neighbourhood when making a decision. The higher the number n_i of "changed" pixels found in this neighbourhood, the lower the threshold is [1]:

$$t(n_i) = T + (4 - n_i) \cdot B, \tag{4}$$

with $0 \leq n_i \leq 8$ when using a 3×3 neighbourhood. The parameter B is a positive–valued potential which determines the range of $t(n_i)$. If $n_i = 0$, the threshold t reaches its maximum value of $T + 4B$. The minimum value of $T - 4B$ results if $n_i = 8$, i.e. all neighbours of pixel i are labelled as "changed". If there are as many "changed" as "unchanged" labels, we have $n_i = 4$, and $t = T$. Clearly, the threshold $t(n_i)$ favours the emergence of compact, smoothly shaped object masks, and reduces scattered decision errors caused by noise.

The labels $q(k)$ necessary to calculate $t(n_i)$ can be obtained as follows: assuming a raster scan from the upper left to the lower right image corner, the labels in the causal part of the 3×3-neighbourhood of pixel i are already known. The labels in the noncausal part of the neighbourhood are approximated by keeping labels from the previous change mask (see Figure 1). Note that when overwriting the previous change mask during the raster scan, this constellation emerges automatically.

[1]Instead of taking the absolute sum, the local squaresum of grey level differences may similarly be used [4]. The absolute sum is, however, more robust against outliers.

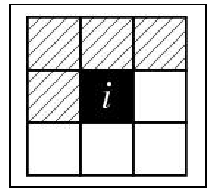

Figure 1. Label constellation in the neighbourhood of pixel i. The causal part of the neighbourhood is shown shaded.

3. Homomorphic change detection

3.1. Homomorphic filtering

The intensity of an image is generated by an incoming illumination, which is reflected by the surfaces of the objects in the observed scene. For Lambertian object surfaces we can model the intensity of the τ–th frame in an image sequence by

$$y_\tau(k) = i_\tau(k) \cdot r_\tau(k), \tag{5}$$

with k being the pixel–index, i the illumination and r the reflectance component [11]. The structure of the depicted scene is captured in the reflectance component r. Consequently, we try to separate i from r and use only the latter for change detection. In many realistic cases the scene illumination can be assumed to be spatially slow–varying, whereas the reflectance component contains also medium high–frequency details, i.e. object information [11]. Therefore we extract the reflectance component by first applying the logarithm and then a linear high–pass filter. The logarithm transforms the multiplicative relation between y, i and r into an additive one, i.e.

$$\log(y_\tau(k)) = \log(i_\tau(k)) + \log(r_\tau(k)). \tag{6}$$

Although the log-nonlinearity modifies the spectral content of illumination and reflectance components, it is in practice often justified to assume the log-illumination to be still spatially slowly varying [11].

Equation 6 holds even if we have to deal with a camera nonlinearity which is often described by the following exponential law (gamma correction):

$$y_\tau(k) = y_{in}^{\gamma}(k) \tag{7}$$

where $y_{in}(k)$ is the detected intensity, and γ the *gamma-value*. The purpose of gamma-correction is to make better use of dynamic range, with γ typically being about 0.4 [8, p. 38]. The multiplicative relation between y, i and r is not distorted as in the log–domain the gamma value is transformed to a gain factor.

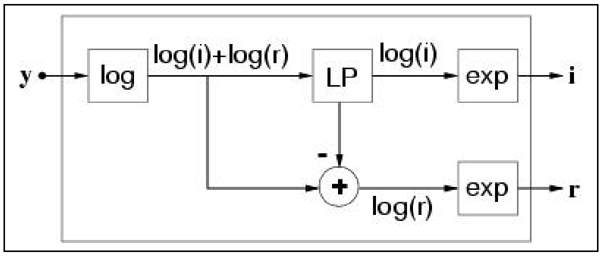

Figure 2. Homomorphic filter for multiplied signals.

The homomorphic filter is shown in Figure 2: After applying the logarithm, the image is low–pass filtered using a binomial filter–kernel and then subtracted from the logarithmic original, yielding a high–pass component. Exponentiation of both high-pass and low–pass components approximately separates the image into illumination and reflectance components. For our purpose, we could omit the exponentiation. However, as the non–linear log–operation makes the camera noise variance σ_u^2 signal–dependent it would be more difficult to handle the parameter σ_{u} in equation 1.

Figure 3 illustrates the effect of homomorphic filtering. The top image is taken from of a sequence with two moving toy engines. Each frame is of size 320×240 pixels. In addition a spot of light crosses the scene quickly from left to right. In the depicted image it is about in the centre of the scene. The middle and bottom images show the reflectance and the illumination components respectively. A binomial lowpass kernel of size 51×51 pixels was used in the homomorphic filter. It is clearly visible that in the reflectance image r illumination effects are strongly suppressed while object information is preserved. In the illumination image i, however, the light–spot is very prominent whereas object details are blurred.

Of course, the illumination image still contains low–frequency parts from the reflectance and thus separation of the two components is only approximate. But in practice the separation works efficiently, as the example in Figure 3 shows.

3.2. Illumination–invariant change detection

To obtain a change detection system which is independent of illumination variations the algorithm described in section 2 is applied to the reflectance components of two successive frames. The full algorithm is depicted in Figure 4. After calculating the reflectance components r_τ and $r_{\tau+1}$ of the corresponding input images y_τ and $y_{\tau+1}$ the change detection is carried out. The sum over the window w_i in equation 1 is implemented by using a moving average low–pass filter (LP) of size 5×5 pixels.

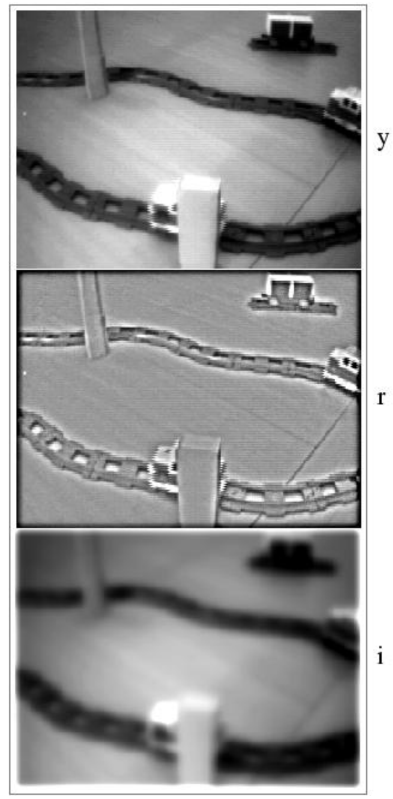

Figure 3. Homomorphic filtering of the image y (top) yields the reflectance r (middle) and the illumination component i (bottom).

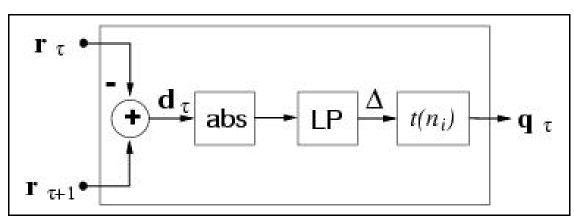

Figure 4. Illumination–invariant change detection.

4. Results

Figure 5 shows two successive frames taken from the sequence with two moving toy engines (top row), where illumination from a moving real light source crosses the scene from left to right. In direct change detection this variation of the scene illumination shows up in the Δ–image (middle left). Therefore, direct application of change detection clearly meddles illumination effects with the desired object mask (middle right). In the bottom row images, the moving light–spot is not visible in the Δ–image, which is now calculated for reflectance images (left). Consequently, the change mask is almost unaffected by illumination changes (right). Hence, the presented illumination insensitive system reacts only to the object motion (plus moving shadow and uncovered background), as desired. The significance level was set to 0.0005 and the cost parameter B to $B = 3.75$ (cf. the values used in [1]). For the reflectance images, σ_u^2 was estimated to $\sigma_u = 8$.

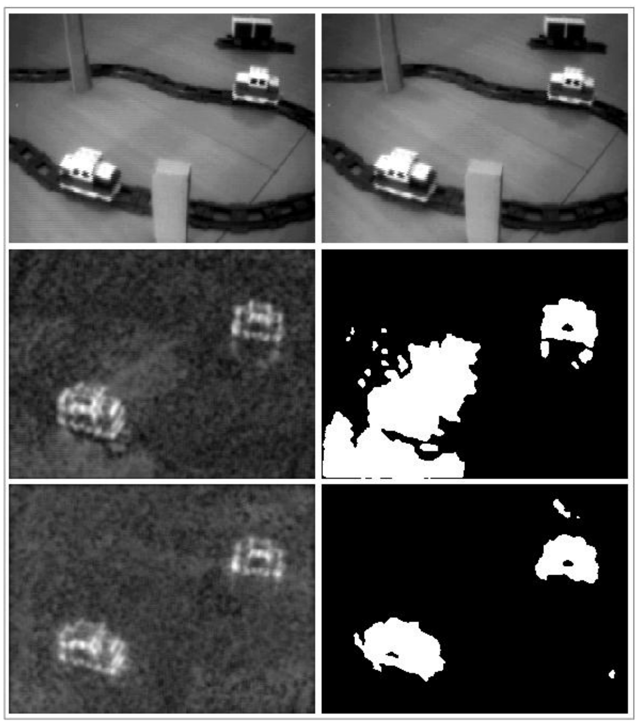

Figure 5. In plain change detection the variation of illumination (top) shows up in the Δ–image and in the change mask (middle row). The images in the bottom row show that the homomorphic system reacts almost exclusively to object motion.

Figure 6 compares the grey level profiles for identical lines in the Δ-images with and without homomorphic prefiltering. The line is marked in white in both pictures, and does not cross a moving object. The observations are hence caused by noise or by the moving light–spot. As the line profiles show, there is a strong ”response” to the moving light–spot (top right) when no homomorphic filter is applied, whereas the Δ-image computed from the reflectance images shows no such response.

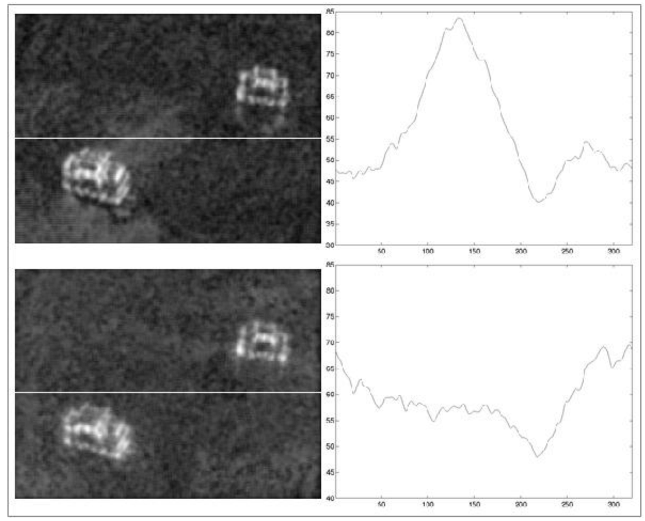

Figure 6. The grey–level values taken from a line in the Δ–images without (top row) and with homomorphic filtering (bottom row).

5. Discussion

In this paper we have developed a new illumination–invariant change detection algorithm by combining the change detection algorithm described in [1] with the homomorphic filter for multiplicative signals from [10, 11]. Clearly, the image model underlying the homomorphic reduction of illumination is only a first-order approximation, which, however, has also been used for other purposese like image enhancement [11] and shadow detection [14]. In our application, results obtained so far from test image sequences with genuinely varying illumination confirm that homomorphic motion detection is insensitive even to fast variations in illumination, without noticeably harming the detection of moving objects. As homomorphic filtering requires only point operations and a separable FIR filter, its computational expense is relatively low. We currently compare our method to alternative approaches (e.g. [9, 13]). Here, the comparison to algorithms in [13] is of special interest, which reduce artifacts from illumination by spatial derivatives or subtraction of the local mean. Both imply additive rather than multiplicative illumination.

So far, the linear lowpass in the homomorphic filter was determined experimentally. A better separation between illumination and reflectance may be obtained by using the stochastic homomorphic filter in [6], which determines the filter parameters based on statistical signal models.

A limitation of our approach is the assumption of spatially slowly-varying illumination. While this assumption is obviously justified in many outdoor conditions, it may not hold for indoor scenes. More elaborate models for illumination identification will then be necessary.

6. Acknowledgement

We are grateful to Marcel Aach for providing the toy engines to record the test sequence used throughout this paper.

References

[1] T. Aach and A. Kaup. Bayesian algorithms for change detection in image sequences using Markov random fields. *Signal Processing: Image Communication*, 7(2):147–160, 1995.

[2] T. Aach, A. Kaup, and R. Mester. A statistical framework for change detection in image sequences. In *Proceedings 13ième Colloque GRETSI*, pages 1149–1152, Juan-Les-Pins, France, September 1991.

[3] T. Aach, A. Kaup, and R. Mester. Change detection in image sequences using Gibbs random fields. In *Proceedings IEEE International Workshop on Intelligent Signal Processing and Communication Systems*, pages 56–61, Sendai, Japan, October 1993.

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

[5] C. Cafforio and F. Rocca. Methods for measuring small displacements of television images. *IEEE Transactions on Information Theory*, 22(5):573–579, 1976.

[6] R. W. Fries and J. W. Modestino. Image enhancement by stochastic homomorphic filtering. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(6):625–637, 1979.

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

[8] B. Jähne. *Digital Image Processing*. Springer, Berlin, 1997.

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

[10] C. R. Moloney. Methods for illumination-independent processing of digital images. In *IEEE Pacific Rim Conference on Coomunications, Computers, and Signal Processing*, pages 811–814. IEEE, Mayy 9–10 1991.

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

[12] B. G. Schunck. The image flow constraint equation. *Computer Vision, Graphics, and Image Processing*, 35:20–46, 1986.

[13] K. Skifstad and R. Jain. Illumination independent change detection for real world image sequences. *Computer Vision, Graphics, and Image Processing*, 46:387–399, 1989.

[14] J. Stauder, R. Mech, and J. Ostermann. Detection of moving cast shadows for object segmentation. *IEEE Transactions on Multimedia*, 1(1):65–76, 1999.