Summary

I Introduction

There are many methods in computer vision to detect objects in image sequences. In this work we want to analyze selected methods for robust object detection even under difficult circumstances. In most cases detection is performed by classification of foreground and background.
In typical detection methods the main distinguishing factor is the *variation of each pixel over time.*
We use a method described by Mittal *et al.* (2009) as basis for our analysis, which provides good results for object detection on fluctuating background.
This method realizes a feature space, where observed parameters describe besides the *time*, the *texture* around the pixel and *texture variations on time* to separate the objects from the changing background.
We analyze and implement this method and compare it with image subtraction methods and with a simple Gaussian method. We evaluate the performance of the methods using different challenging image sequences (Fig. 1).

Figure 1: Examples for dynamic Scenes

In Chapter II we give a short overview on related work. Chapter III presents the implemented methods of object detection. Finally in chapter IV we present the results and discuss them.

II Related Work

Image subtraction methods are a simple and often used to determine movements in image sequences with static background. Generally, a suitable reference image is subtracted from the current image, to get information about the object motion. Further information about image subtraction methods can be found in Nischwitz & Fischer (2007) P.597ff. and Jähne (2002) P.395ff. and P.541ff.
The literature provides a wide variety of possibilities, to implement object detection by using Gaussian methods.
For example, Elgammal *et al.* (2000) present a method that uses a combination of K distributions instead of one distribution. A background pixel that represents sometimes

a leaf and sometimes an object behind the tree can already be classified as background pixel with a mixed distribution. The method of Elgammal *et al.* (2000) is based on a non-parametric density model.

Friedman & Russell (1997) are using an EM algorithm for the detection of moving cars on a highway. They estimate Gaussian distributions to model the visual properties of background and foreground in scenes with street traffic. The data of the image sequence is classified in `road`, `shadow` and `vehicles`.

These are typical methods used to detect objects in image sequences. All of them are using gray values or functions of gray values as features to separate the foreground from the background. Now let us consider a new approach based primarily on the work of Mittal *et al.* (2009) . They use an alternative approach to model the background and to detect changes in images. We want to distinguish the structured movements of the background from those of the object. For this we determine detection parameters, wich take variations of the pixels in texture, spatial and temporal level into account.

III Detection Methods

We have developed a GUI[1] in which we can choose between the different methods, their parameters and the datasets. In the following, we present the used methods.

Subtracted Images:

We search the subtracted image. The reference image can be represented by a static background image (e.g. the first image in the sequence), a mean image determined out of the last images or by a previous image.

Using Gaussian distributions:

In a simple method, we look for the variation of the pixels over time. We estimate for example the Gaussian distribution of each pixel in each image, based on the last T images. We determine an arithmetic mean μ_t and an empirical variance σ_t^2. Now we can compare the current gray value $g(i,j)$ of the pixel with the resulting Gaussian distribution in a two-sided test. If the value is within the distribution $N(\mu, \sigma)$ with a certain probability, we assume that this pixel represents the background. If the value is outside of the acceptance range, we detect foreground.

This method works reliable, as long as the objects perform fast movements and as long as their trajectories are relatively homogeneous. The background may vary only slightly. A background in which the intensity of the gray levels vary widely leads to poor detection rate.

Mittalmethod:

We introduce two detection parameters. Basis of the procedure is an alternative and compact feature space. The detection is not pixel accurate. It is performed for small block regions. For each image block we calculate an average. We subtract the average image from the actual image to get the reduced image. Then we vectorize this reduced image for each block and each time. Out of the last T vectored images we calculate a covariancve matrix.

[1]General User Interface

With this matrix we calculate a PCA[2]. With the convolution of each image with each filter of the resulting linear filter bank, we determine the significant movements. Based on this result we can estimate a predicted image and can calculate the two detection parameters. These parameters are basically so called Mahalanobis-Distances. The first detection parameter detects the *structural changes* in the block, the second one detects *changes in motion characteristics*. If both parameters pass the chi-square test, we detect foreground. If the algorithm is successfully passed, we get for every block and every time a detection decision (Fig. 2).

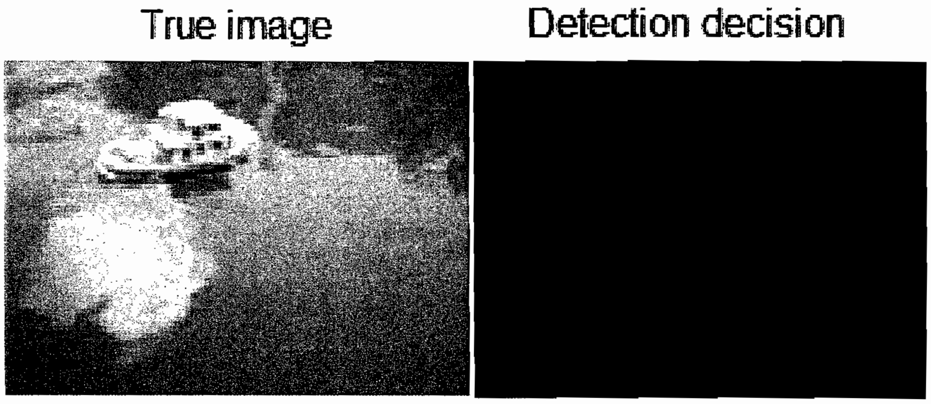

Figure 2: True image and resulting detection decision (Mittalmethod).

Mittal using non-linear filters:
We change the method of Mittal *et al.* (2009) further by creating various non-linear filter banks instead of a linear filter bank (e.g. Fig. 3). This should lead to higher information content if we allow a larger number of principal components in our PCA.
We change our input data. Instead of one image, we use several filtered images. Our input was one image before, and now our input is given by as many images as filters in the filter bank. This leads to correspondingly higher computational complexity. Formation of the covariance matrix and reduction of the feature space require higher computational effort. Using the nonlinear filter bank, we can detect more complex movements.

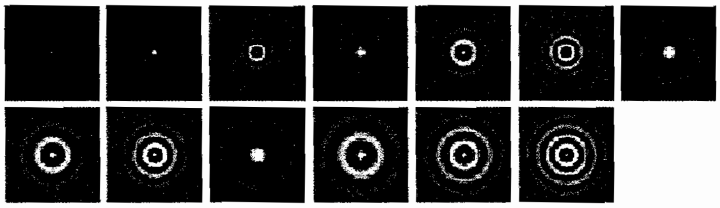

Figure 3: Schmid-Filterbank.

IV Results and conclusions

We evaluated the methods with different data sets. We could show that especially in scenes with fluctuating background the method of Mittal *et al.* (2009) leads to good detection,

[2]Prinicpal Component Analysis

because we consider in this method time and texture, and texture variation over time in the sequence.

Using a statistical test instead of setting a threshold appears usefull. Making a detection decision based on a chi-square test leads to a clear conclusion, where we detect foreground. The change of the input information with regard to alternative nonlinear filter banks led to slightly improved results. Often the block size can be reduced without degrading the detection quality.

We have noted, that the Mittal-method still works, even if Gaussian methods already fail. The two parameters of this method are able to detect salient motion in front of fluctuating background. Figure 4 shows an example for these detection values. Figure 5 compares the methods with each other by showing some example images.

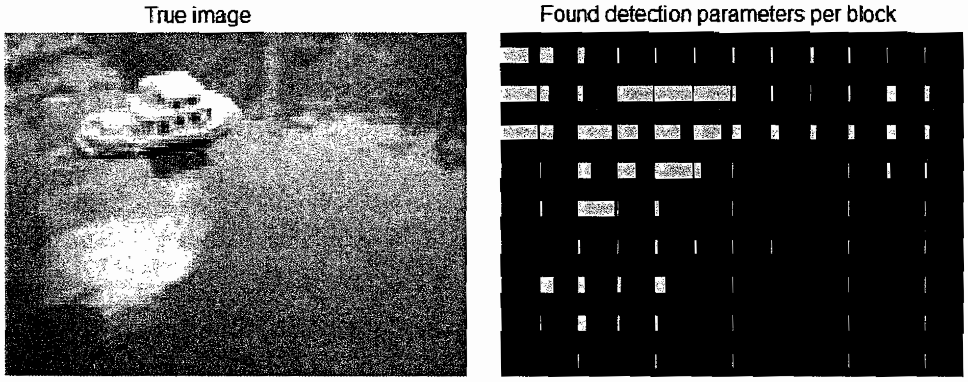

Figure 4: Detection parameters identified in a pool scene image. Structural change is represented by dark gray, change in motion characteristics by slightly gray.

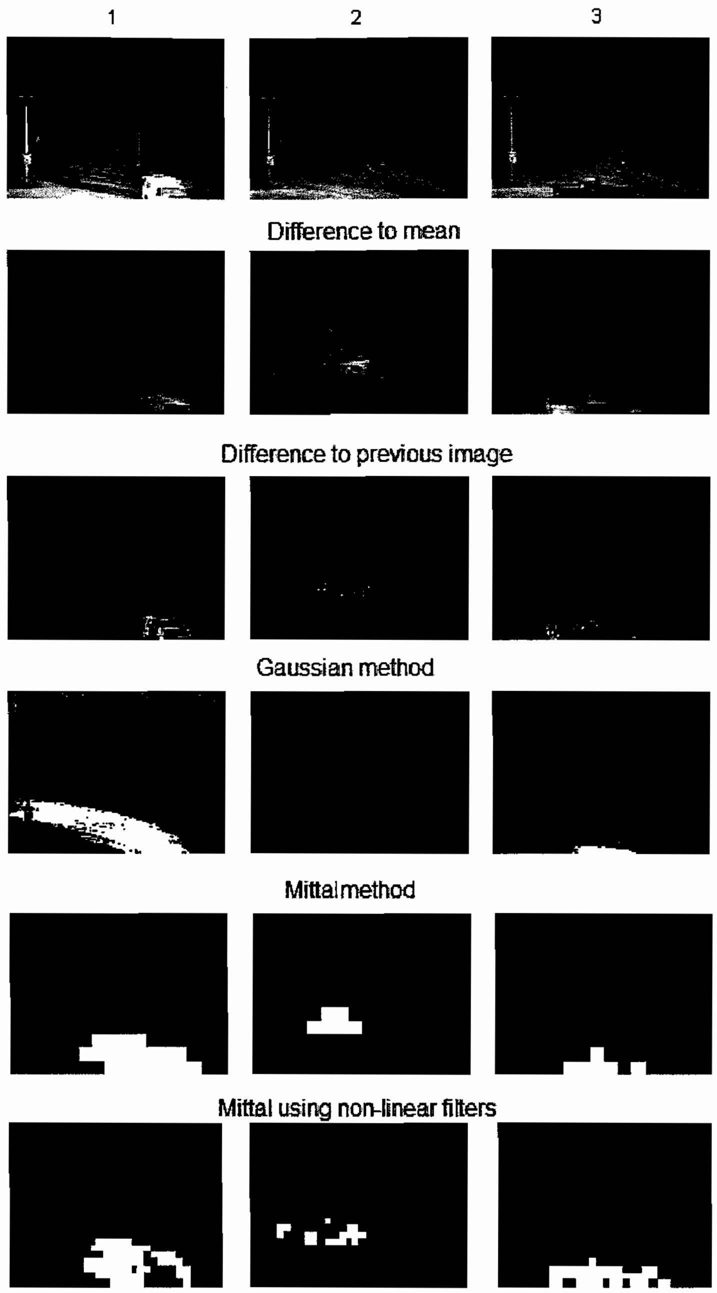

Figure 5: Results of the implemented detection methods. Only the Mittalmethod and the extension with non-linear Filters can detect the black pick-up in the background.