UNIVERSITEIT
VAN
AMSTERDAM

Global intensity correction in dynamic scenes

P.J. Withagen[1,], **K. Schutte**[2], **and F.C.A. Groen**[3]
[1]X-ray, BL Components
Philips Medical Systems
Paul.Withagen@philips.com
[2]Electro Optics group
TNO Defence, Security and Safety
Klamer.Schutte@tno.nl
[3]Intelligent Autonomous Systems group
University of Amsterdam
F.C.A.Groen@uva.nl

Changing image intensities causes problems for many computer vision applications operating in unconstrained environments. We propose generally applicable algorithms to correct for global differences in intensity between images recorded with a static or slowly moving camera, regardless of the cause of intensity variation. The proposed intensity correction is based on intensity-quotient estimation. Various intensity estimation methods are compared. Usability is evaluated with background classification as example application. For this application we introduced the PIPE error measure evaluating performance and robustness to parameter setting. Our approach retains local intensity information, is always operational and can cope with fast changes in intensity. We show that for intensity estimation, outlier removal is essential for dynamic scenes. For image sequences with changing intensity, the best performing algorithm (MofQ) improves foreground-background classification results up to a factor two to four on real data.
Keywords: Intensity correction, Intensity variation, Camera calibration, Pixel classification, Time-varying imagery

IAS

intelligent autonomous systems

Contents

Intelligent Autonomous Systems
Informatics Institute, Faculty of Science
University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam
The Netherlands

Tel (fax): +31 20 525 7461 (7490)
http://www.science.uva.nl/research/ias/

Corresponding author:

F.C.A. Groen
tel: +31 20 525 7461
Groen@science.uva.nl
http://www.science.uva.nl/~Groen/

1   Introduction

Computer vision has proved very successful in well-constrained industrial environments (for instance when illumination, objects types, and orientations are known). However, in many practical applications, including airborne or remote sensing, medical imaging, face recognition, outdoor robotics, and surveillance applications, the environment can hardly be controlled. Illumination changes by lights switching on or off, or by clouds moving in front of the sun. Automatic gain control (AGC), white balance and iris are often applied to optimally map the amount of reflected light to the digitizer dynamic range. However, when scene content changes they cause changing image intensity over time.

Problems then arise with many algorithms that assume Constant Image Brightness (CIB) or that are based on the Brightness Constancy Constraint Equation (BCCE). Applications where image intensity changes cause problems include background subtraction, object tracking, stereo matching, optical flow computation, video coding and image retrieval.

In this paper we introduce algorithms for correction of global intensity changes in image sequences. The algorithms estimate and correct for global temporal intensity variations. They can be applied as pre-processing step for other image processing algorithms as mentioned above. We limit ourselves to closed form solutions to guarantee possible implementation for real-time applications. Based on a model of a CCD camera a number of algorithms are proposed. The algorithms are evaluated on both simulated and real images.

Evaluation of the algorithms is performed on three criteria. First, the precision of the parameter estimation is evaluated using the sum of squared differences between a reference image and the corrected image. Second, the usability of the correction is evaluated using the performance of a representative post processing algorithm. Third, the robustness of the post processing is evaluated. For all evaluation methods we will show the impact of outlier removal.

For real-world application we will focus on the foreground-background classification problem. We use the popular online Expectation Maximization (EM) algorithm to estimate a multi-Gaussian model of the background color for each pixel, see [18, 22, 28]. The model is used to classify pixels as foreground or background. This is an essential step in many surveillance applications.

This paper is structured as follows: in section 2 we discuss existing techniques to handle changes in intensity. We will introduce our model of changing intensity in section 3. There we will consider the differences between changes in intensity caused by the combination of a changing scene and automatic gain control, and by changing illumination. In section 4 the proposed methods are described, according to a model of the CCD camera evaluated in Appendix A. These methods are evaluated by both simulation and experiments on real images in section 5. Finally, conclusions are presented in section 6.

2   Previous Work

There exists an extensive amount of literature concerning applications like moving object detection, stereo matching or optic flow calculation. Considering that these algorithms normally expect constant image intensity, it is surprising how little work is done in the area of intensity correction. In this section we shortly introduce the different techniques that are available for intensity correction. We evaluate them on accuracy, usability and computational complexity.

The intended algorithm will be used for dynamic scenes with moving objects. These objects cause a scene change and can decrease the accuracy of the intensity correction. It is therefore important to have outlier removal, as will be shown in the experiments in section 5. The simplest way of dealing with changes in intensity is ignoring all intensity information. Intensity invariants [21] or normalized colors can be used. Instead of intensity information other features can be used,

for example color [3, 9], edges [11, 13], or depth [5]. However, using these features disregards the information available in the intensity.

Local changes in intensity are beyond the scope of this paper. However, techniques dealing with local changes could be used to correct for global changes in intensity. For example, the pixel ordering in a neighborhood [29] can be used, one can use a maximum likelihood estimate in an image region [16], or one can use the low frequencies [23]. For background modelling, one can consider shadow detection, see for example [27, 9, 10, 17]. The use of Kalman prediction to the kernels in the EM model [22] is another example of a local technique in conjunction with background modelling. The drawback of these techniques is that they ignore the fact that all pixels change simultaneously. This leads to a less accurate estimate of the global effect.

Literature reports complicated methods: dynamic histogram warping changes image intensities such that the histograms of the two images become equal [2], by an iteratively weighted least squares estimation, the bias, gain and gamma of an image can be estimated and corrected for [25], and estimation of the gain and bias together with optic flow allows the use of optic flow under global intensity changes [1]. Drawback of these methods is the high computational complexity. This makes real-time implementation difficult and expensive. Also, [2] and [1] do not perform outlier removal.

Considering background modelling techniques, a frequently used approach for dealing with changes in image intensity is relying on the adaptation speed of the background modelling technique [15]. However, this adaptation will only resolve the problem for relatively slow changes in intensity, and it is difficult to tune the update speed of the model. A high update speed may learn slow objects into the background model (missed objects), while a low update speed can be unable to adapt the model fast enough to cope with changes in intensity (false alarms), see [24]. No single update speed guarantees acceptable results for all possible situations.

An approach, closely related to relying on the adaptation of the background classification algorithm, is setting a limit on the fraction of allowed foreground pixels (e.g. 70 %). When this fraction is exceeded another background model (if available) is chosen or the background model is re-initialized [13, 24]. The performance will decrease in applications where illumination or gain changes frequently occur, because after re-initialization a new background model must be learned. During each learning period classification results are unreliable.

The idea of using multiple instances of a scene taken under different illumination conditions is also considered in [4, 8]. They use an eigenspace method to overcome complex changes in illumination (not only intensity but also illumination direction may change). An interesting approach in conjunction with image retrieval is given in [12]. Their method assumes that the pixel-wise image ratio of two images from the same object is simpler than the ratio of two different objects (they define simplicity as based on the complexity of the algebraic function needed to locally approximate the shape of the image). This assumption allows for object comparison under complex changes in illumination. Only one measure of similarity is calculated for the entire image. Therefore, usability is restricted to applications like face recognition and image retrieval, so this is less general than the scope of this paper, were we are aiming at a generally applicable method.

A useful approach for real-time applications is the direct calculation of the intensity difference between two images. This is done using the average [6] or a least squares estimate [14]. However, these methods are sensible to outliers.

In this paper we want to develop a method for the correction of global changes in intensity for a static or slowly moving camera that overcomes the above limitations. The method should be generally applicable and insensitive to outliers in dynamic scenes. It should make optimal use of local intensity information. Furthermore, it should have low computational cost so that it can be implemented in real-time.

3   Model Description

In this section we present a correction algorithm for global changes in intensity. We then introduce a simplified model of a CCD camera. This model has been experimentally verified for a range of cameras in Appendix A. Using this model we then introduce the apparent gain factor that needs to be estimated.

3.1   Global intensity correction

The goal of this paper is to correct for global differences in intensity between two images. Consider two images i_r and i_t depicting an equal scene at different time instances. There is global difference in intensity between the two images. We intend to correct for this intensity difference by

$$i_{t,\text{corrected}} = \frac{i_t}{a} \ , \tag{1}$$

with a the apparent gain factor. $i_{t,\text{corrected}}$ and i_r have equal global intensity. We will give an equation for a in this paper.

3.2   Model of CCD cameras

Based on a general model of CCD cameras [7], we report in Appendix A the experimental results of the different contributions in the CCD model with added gamma correction. An important conclusion from this work, also published in [26], is that it has been shown that: "It is wrong to pick a general model and assume its validity. It is important to validate the model for the specific camera used."

For the cameras used in Appendix A that do adhere to the model, the experimental results allow for simplification of the model. For sufficiently large intensity values, both offset and additive noise can be neglected. The simplified model of the CCD camera to be used in the remainder of this paper is now given by

$$i_t = g_t^\gamma \left(h_t i_0 + N_{\text{S}} \right)^\gamma \ , \tag{2}$$

with g_t the camera gain, i_0 the scene irradiance, h_t a factor related to the camera shutter time, iris size and scene illumination, N_{S} multiplicative noise and γ the value of the gamma function.

3.3   The apparent gain factor

The image recorded at time t will be compared to some reference image r recorded earlier. These images are given by

$$i_t = g_t^\gamma \left(h_t i_0 + N_{\text{S}} \right)^\gamma \tag{3}$$

$$i_r = g_r^\gamma \left(h_r i_0 + N_{\text{S}} \right)^\gamma \ . \tag{4}$$

Changes in image intensity can be caused by either changes in the camera gain g_t or changes in the illumination intensity, iris or shutter resulting in a changed h_t.

Considering equal scene, the relation between i_t and i_r is

$$i_t = a i_r + N_{\text{total}} \ , \tag{5}$$

with a equal to that defined in equation 1. As noise N_{total} is zero-mean, this shows that equation 1 can be used for global intensity correction.

The apparent gain factor a is given by:

$$a = \left(\frac{g_t h_t}{g_r h_r}\right)^{\gamma} \ . \tag{6}$$

Considering the noise of the two images independent gives for the variance in N_{total}:

$$\sigma_{\text{total}}^2 \approx \left((a^2 g_r^2 h_r + g_t^2 h_t) i_0 \sigma_t^2\right)^{\gamma} \ , \tag{7}$$

with σ_t^2 he variance in the noise in each of the images.

Regardless of the cause, the effect on the apparent gain factor a is the same, see equation 6. Therefore, without loss of generality we can limit ourselves to the estimation of a in the remainder of this paper. However, there is a difference in effect on the image noise. When the change in intensity is caused by camera gain the increase in noise is different than when the illumination intensity, iris or shutter time changes[1].

4 Proposed algorithms

To correct for changing intensity we need to estimate the apparent gain factor a between two images i_r and i_t defined in subsection 3.1. To obtain constant intensity over time we compare all images to the reference image i_r. This results in the apparent gain factor between the reference image and each other image. These apparent gain factors will be used to correct the corresponding images.

Because of the noise in both images, this parameter estimation problem is not trivial. Besides a theoretically optimal estimate, several alternative algorithms will be given here. They will be compared experimentally in section 5.

4.1 Estimation of the intensity factor

We intend to give an accurate non-iterative estimate of the apparent gain factor a. The less computation an algorithm requires, the easier it can be implemented in real-time. Therefore we also present simplifications to the theoretically optimal algorithm. In section 5 we will experimentally compare these algorithms.

As the majority of the noise contributions are multiplicative, a Weighted Least Squares (WLS) estimate is a theoretically optimal non-iterative estimate. Minimizing the criterion

$$L_2 = \sum_{s \in S} w_s^2 (i_t - a i_r)^2 \ , \tag{8}$$

for all pixels s in set S in least squares sense gives

$$a_{\text{WLS}} = \frac{\sum_{s \in S} w_s^2 i_{r,s} i_{t,s}}{\sum_{s \in S} w_s^2 i_{r,s}^2} \ . \tag{9}$$

The weights w can be calculated using the inverse of the image noise, which depends on the intensity. Using the common simplification of only multiplicative (shot) noise leads to the following weights

$$w_s^2 = \frac{i_{r,s}}{i_{t,s} i_{r,s} + i_{t,s}^2} \ , \tag{10}$$

---

[1] For an equal change in intensity, the variance in the noise $\sigma_{\text{total},g}^2$ caused by changing camera gain is approximately $\sigma_{\text{total},g}^2 = \frac{1+n}{2} \sigma_{\text{total},h}^2$, for $\gamma = 0.5$ and with n the amount of change $n = \frac{g_t}{g_r}$ or $n = \frac{h_t}{h_r}$ respectively.

using the approximation $a = \frac{i_{t,s}}{i_{r,s}}$ in the weights. This will however increase the influence of low intensity values. For these values the simplification used is not valid, as the amount of additive noise will be larger than the amount of multiplicative noise.

To reduce the influence of low intensities and at the same time reduce computational requirements, we can use equal weights. This gives us the standard Least Squares (LS) estimate

$$a_{\mathrm{LS}} = \frac{\sum_{s \in S} i_{r,s} i_{t,s}}{\sum_{s \in S} i_{r,s}^2} \ . \tag{11}$$

Even simpler and requiring fewer computations is using only the Quotient of the Average (QofA), related to the L_1 criterion:

$$a_{\mathrm{QofA}} = \frac{\sum_{s \in S} i_{t,s}}{\sum_{s \in S} i_{r,s}} \ . \tag{12}$$

All methods given above calculate the quotient of two numbers. For statistical outlier removal, see subsection 4.2, it would be profitable to have an intensity ratio per pixel. This can also be useful for the extension to local intensity correction, see [27] Therefore, we also take the Average of the pixel-wise Quotient (AofQ) into account

$$a_{\mathrm{AofQ}} = \frac{1}{|S|} \sum_{s \in S} \frac{i_{t,s}}{i_{r,s}} \ , \tag{13}$$

with $|S|$ the number of pixels in the set of pixels S.

The experiments will show that it is important to perform outlier removal. Also, the quotient between two images typically has a positively skewed distribution where the mean would overestimated the apparent gain factor. For these reasons the median is also taken into account

$$a_{\mathrm{QofM}} = \frac{\mathrm{M}_{s \in S} i_{t,s}}{\mathrm{M}_{s \in S} i_{r,s}} \ , \tag{14}$$

and

$$a_{\mathrm{MofQ}} = \mathrm{M}_{s \in S} \frac{i_{t,s}}{i_{r,s}} \ , \tag{15}$$

where M denotes the median of a set of numbers.

4.2 Outlier Removal Algorithm

Comparing the two images i_r and i_t we should take into account that not all pixels will be stationary in dynamic scenes. Pixels depicting dynamic scene violate the assumption of equal scene introduced in subsection 3.1. The equations given above are thus only valid for pixels depicting stationary scene. Non-stationary pixels should be excluded from the apparent gain factor estimation.

We will investigate outlier removal based on statistics. For applications using foreground/background classification, moving objects will cause outliers. In such case classification between foreground and background can be used for outlier removal. However, this makes the global intensity correction algorithm less general.

Statistical outlier removal is based on the observation that the majority of the pixels depict the same scene in both images. We calculate the average μ_q and standard deviation σ_q of the pixel-wise ratio q_s between the two images. Pixels for which $|q_s - \mu_q| < T_{\mathrm{outlier}} \sigma_q$ holds are labelled as outlier. T_{outlier} should be chosen such that enough pixels remain for a statistically accurate estimate, but that only those pixels remain that are very unlikely to be outliers. We will use $T_{\mathrm{outlier}} = 1$.

Another set of pixels that should not be taken into account are pixels that are close to either the upper or lower bound of the range of pixel values. For pixels close to the lower bound the additive noise and dark current cannot be neglected, while pixels close to the upper bound may suffer from saturation problems. Therefore, we will ignore pixels within 10 percent of the upper and lower bound.

4.3   Creation of the Reference Image

All algorithms proposed need a reference image to compare the current image to. In our experiments, we will use the first image of the image sequences as reference image. This is the most general solution and requires the least amount of computations. For most applications however, the reference image should be updated as the scene might change. Some alternatives will be discussed below.

The reference image can be periodically renewed by selecting a new image from the input. The image can be selected at random, or based on the amount of background it contains. The latter is preferred as it will contain less moving objects.

When using background modelling [18, 22] it is possible to use the background model as reference. With a Gaussian mixture model, the average of the kernel with the largest weight can be used to compare the current image to, but it is also possible to use the mean of the best fitting kernel. The latter is expected to give better results for multi-modal backgrounds. Compared to choosing an image from the sequence, we expect the amount of noise and the number of foreground pixels in the EM–model reference image to be lower. An additional advantage of using the EM–model to create the reference image is that it is always up-to-date. When the background changes, the reference image is automatically adapted. Drawbacks are a small amount of additional computation.

For a slowly moving camera, each image can be compared to the previous image. If the motion is (approximately) known, only the overlapping area between the images should be used for apparent gain factor estimation.

5   Experimental Evaluation

We will experimentally evaluate the proposed algorithms. We will use simulated images to evaluate the accuracy and usability of the different intensity estimation algorithms in subsection 5.1. We compare the algorithms to each other and to ground truth. Also, the need for outlier removal is evaluated. The effect of intensity correction in conjunction with classification between foreground and background on real images is demonstrated in subsection 5.2. We look into the runtime of the algorithms in subsection 5.3. A discussion of experimental results is given in subsection 5.4.

For the evaluation of the usability we demonstrate intensity correction in conjunction with foreground/background classification. For each image the apparent gain factor is calculated and the image is corrected for it. With this corrected image, a model of the background is updated using the online Expectation Maximization (EM) algorithm [18]. Four kernels are used to model the background. They are updated with an update speed u_B.

The classification algorithm proposed by Stauffer [22] is used to do classification between foreground and background. It is based on the assumption that the color distributions of the foreground and background are distinct. This way, separate kernels will be used to model the foreground and background. Each kernel is assigned a label, determining whether it models foreground or background. The background kernels together describe more then a fraction F of all data. We use $F = 0.5$ as proposed by Stauffer. Pixel classification is performed by determining whether a pixel can be assigned to any of the kernels labelled as background. A

pixel is assigned to a kernel if it's value is within T_{Stauffer} times the kernels standard deviation from the kernel mean (Stauffer proposes $T_{\text{Stauffer}} = 2.5$ in [22]).

5.1   Comparison by simulation

We will evaluate the proposed algorithms using a simulated image sequence. For each image the apparent gain factor a is estimated and the image is corrected according to equation 1. The corrected image is used to measure the performance of the intensity correction algorithm and to perform foreground/background classification.

5.1.1   Image generation

Images are generated based on the full model of the CCD given by equation 20, so without simplifications. As scene we use the red channel of the first image from the Intratuin image sequence, see figure 3. In each frame 20 % of the pixels is selected at random and they are given a foreground value. The foreground value is drawn from a uniform distribution between 0 and 0.5, while the distribution of the background lies between 0 and 1 with peaks around 0.3, 0.6 and 0.8, see figure 5.1.1 for a noiseless example image and figure 5.1.1 for its histogram. We choose the parameter settings of the image simulation such that they approach an average camera from the cameras characterization in Appendix A.

The scene intensity is kept constant $h_t = h_r$ and the camera gain g_t is varied, see figure 2. The camera gain is 1 for images 1 to 150 for training and evaluation of the algorithms on images without changes in intensity. The intensity linearly decreases from 1 at image 150 to 0.5 at image 200 for the evaluation of the algorithms on images with a changing intensity.

Besides the image sequence i_t, two additional images are created. The reference image i_r which is used for the estimation of the apparent gain factor together with the current image, and the ground truth image i_{gt}. These images have equal illumination and camera gain and do not have foreground pixels. The ground truth image is noiseless, the reference image contains noise.

(a) Simulation image
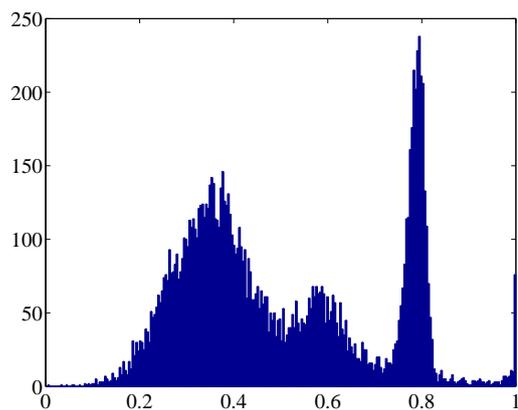
(b) Histogram of simulation image

Figure 1: Simulation image and its histogram. The image on the left shows an example image used for the simulation experiments. On the right the histogram of this image is given.
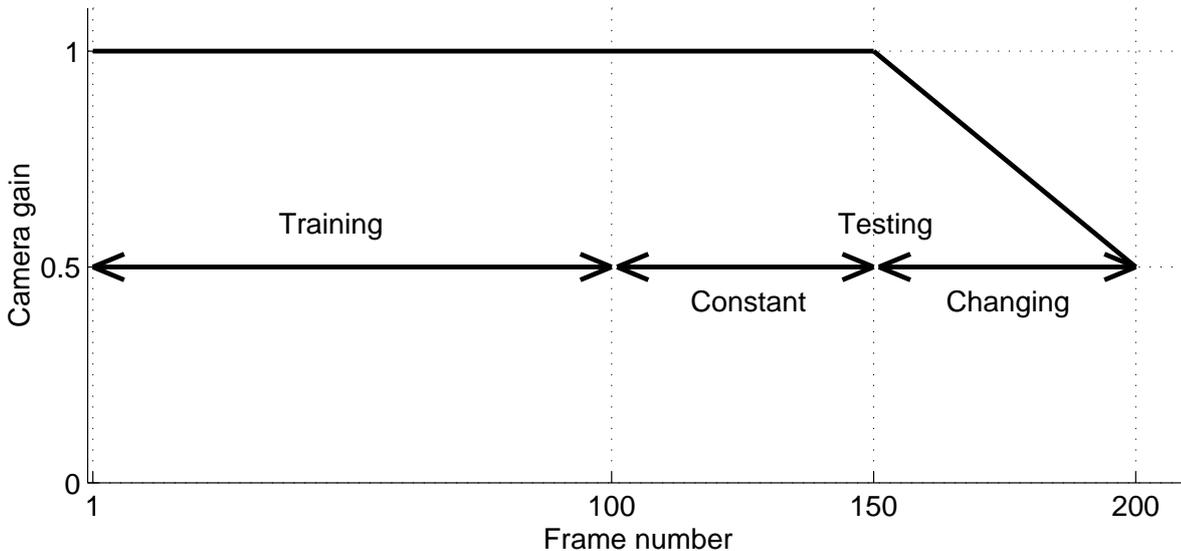
Figure 2: The value of the camera gain for the simulated images.

5.1.2 Evaluation criteria

Two criteria for evaluation are used, one for the accuracy of the apparent gain factor and one for the usability in conjunction with foreground/background classification.

For accuracy evaluation we use the average root mean squares error between the intensity corrected current image and the ground truth image for all pixels depicting background:

$$e_{\text{accuracy}} = \frac{1}{|B|} \sum_{b \in B} \left(i_{\text{gt},b} - \frac{1}{a} i_{t,b} \right)^2 , \tag{16}$$

with $|B|$ the number of pixels in the set of background pixels B. Note that as we use a realistic setting with noise in the current image, this criterion gives for perfect intensity correction the standard deviation of the image noise.

The usability of the intensity correction algorithm is evaluated with foreground/background classification. A Mixture of Gaussians models is updated using online Expectation Maximization, with an update speed $u_{\text{B}} = 0.05$. With this model, classification is performed using the algorithm of Stauffer with a threshold $T_{\text{Stauffer}} = 3$. These settings were found to be optimal for the reference method, without intensity correction, and the given evaluation criterion for usability defined below. Evaluation is started after 100 frames, allowing the EM background model to learn on the first 100 frames.

The impact of the intensity correction for our application is expressed in a usability measure: the classification performance. It shows wether using intensity correction gives an improvement in the results. We define the usability of the intensity correction as the fraction of erroneously classified pixels

$$e_{\text{usability}} = \frac{N_{\text{erroneous}}}{N_{\text{total}}} , \tag{17}$$

with $N_{\text{erroneous}}$ the number of erroneously classified pixels and N_{total} the total number of pixels. This assumes equal cost for missed foreground pixels and erroneously detected foreground pixels.

5.1.3   Simulation results

Besides the six proposed methods, correcting with the ground truth a (GT) and using the uncorrected image (No) are also presented. The different algorithms for estimating the apparent gain factor are evaluated with and without outlier removal. For all results, the mean and standard deviation over the images were calculated for the two sections: static and changing intensity. The results shown are averaged over fourteen independent noise realizations.

The results for all combinations are given in tables 1 and 2. Most important observations are a ten times lower accuracy error and a three times lower usability error for methods using outlier correction compared to no correction. Methods without outlier correction perform significantly worse.

From the accuracy results we see that as long as there is some kind of outlier removal (statistical or methods based on the median), accuracy is for most combinations close to the image noise. Exception is the method QofM, which performs slightly worse. For changing intensity, the error with correction is ten times lower than without correction. Without outlier removal results are significantly worse, even on the section with constant intensity.

For usability the results are even more consistent. All combinations that do some kind of outlier removal (statistical or median) give errors close to the error of correcting with the ground truth. The error of 6.8% is caused by the fraction of the 20% foreground pixels that overlap in color with the scene. This should be compared to no correction, where the error triples in dynamic situations. Combinations without outlier removal perform again worse than those with.

As could be expected the WLS method is only optimal when its assumptions are fulfilled, in particular with respect to the probability density distributions, so without outliers. The method is extremely sensitive to outliers, shown by the results without outlier removal. Even the few outliers remaining after outlier removal are sufficient to decrease the performance of the WLS method. To a lesser extend, the same holds for LS. A solution would be to use robust iterative estimators, like M–estimators, but those are very time-consuming and difficult, if not impossible, to use in real-time. The difference in performance between LS and WLS suggests that an additive noise model (LS) gives in this case a better description of the data than the multiplicative noise model used in WLS.

5.2   Evaluation using real image sequences

In simulation experiments optimal parameter settings for the background classification algorithm can be used. In practical situations these optimal settings are often unknown, and the

Table 1: Accuracy results for simulated data. Shown is the average of the error e_{accuracy} in percentages (lower is better). The standard deviations over the frames are given between brackets.

	Static intensity		Changing intensity	
Outl. rem.:	No	Yes	No	Yes
Method:				
No	1.261 (0.009)		12.56 (7.28)	
GT	1.261 (0.009)		1.2660 (0.008)	
QofA	2.640 (0.094)	1.2628 (0.009)	2.631 (0.076)	1.268 (0.008)
QofM	2.402 (0.183)	1.2854 (0.038)	2.319 (0.169)	1.307 (0.047)
AofQ	2.361 (0.083)	1.2619 (0.009)	2.361 (0.072)	1.267 (0.008)
MofQ	1.263 (0.011)	1.2606 (0.009)	1.277 (0.008)	1.266 (0.008)
LS	2.897 (0.107)	1.2637 (0.009)	2.878 (0.084)	1.269 (0.008)
WLS	7.496 (0.410)	1.2714 (0.009)	7.199 (0.354)	1.276 (0.009)

<table>
<tr><td>(a) 500</td><td>(b) 750</td><td>(c) 1000</td><td>(d) 1250</td></tr>
</table>

Figure 3: Some images from the Intratuin sequence.

system will operate in general at sub-optimal settings of these parameters. So the sensitivity of the performance to suboptimal settings will play a major role in the evaluation on real image sequences.

5.2.1 Image sequences

Three image sequences were used for the evaluation, see also figures 3, 4 and 5:

- *Intratuin*: Parking lot with waving tree branches. In this sequence there is no significant variation in intensity. The sequence contains cars and pedestrians, moving both slowly and fast. This sequence has 150x350 pixels and 1250 frames.

- *Schiphol*: Recorded in the main hall of Schiphol airport. There are a lot of global intensity variations due to automatic gain control. The sequence contains relatively large objects, some of which become stationary. This sequence has 90x120 pixels and 1750 frames.

- *PETS 2001*: We used a cut-out from the images of dataset 3, training, camera 1 from the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance 2001 (PETS). There are large local changes in illumination intensity due to clouds. The images contain relatively few object pixels. The part of the image that is used is between rows 300 and 520, skipping the odd rows and between columns 350 and 750, skipping the odd columns. This sequence has 120x200 pixels and 5500 frames.

All sequences are RGB color video data with eight bit per color. For each sequence, five to eleven images were manually labeled. Each pixel was labeled: foreground, background or any. The label any is used for the edges of objects, where it is difficult (for a human) to decide whether this pixel should be labeled as either foreground or background. It is also used for some artifacts in the images like moving objects that stop moving.

The Intratuin and Schiphol sequences are recorded by us and they are available through our website (www...(left blank for review)), together with the manually labeled ground truth for all sequences. The PETS 2001 data is available through pets2001.visualsurveillance.org.

5.2.2 ROC performance

We evaluate the intensity correction algorithms based on their usability in conjunction with classification between foreground and background. The images are corrected with each of the proposed intensity correction algorithms after which the background model is updated and foreground/background classification is performed.

The choice of the best algorithm for foreground/background classification depends on the application. In order to make a good choice, the ratio of the cost of a false alarm and the cost of a missed detection must be known. For the experiments on simulated images described in subsection 5.1, unity cost was assumed.

Table 2: Usability results for simulated data. Shown is the average of the error $e_{\text{usability}}$ in percentages (lower is better). The standard deviations over the frames are given between brackets. Note that these results cannot be compared to the results on real data as only one color has been used in the simulation.

	Static intensity		Changing intensity	
Outl. rem.:	No	Yes	No	Yes
Method:				
No	6.822 (0.716)		17.87 (5.16)	
GT	6.822 (0.716)		6.699 (0.664)	
QofA	8.096 (1.162)	6.813 (0.708)	6.887 (0.713)	6.706 (0.665)
QofM	7.894 (1.066)	6.876 (0.694)	6.873 (0.716)	6.760 (0.669)
AofQ	7.751 (1.027)	6.817 (0.711)	6.828 (0.706)	6.705 (0.665)
MofQ	6.823 (0.717)	6.822 (0.716)	6.710 (0.652)	6.710 (0.661)
LS	8.403 (1.335)	6.819 (0.705)	6.937 (0.714)	6.705 (0.664)
WLS	18.11 (6.38)	6.820 (0.708)	7.927 (1.102)	6.709 (0.664)

(a) 250   (b) 750   (c) 1250   (d) 1750

Figure 4: Some images from the Schiphol sequence.

(a) 2000   (b) 3500   (c) 4500   (d) 5500

Figure 5: Some images from the PETS sequence.

When comparing different classification algorithms, the parameter settings of all algorithms should be optimized for the chosen cost ratio. For a different application, a different ratio might be in use and therefore a different algorithm might be optimal. It would be efficient to compare for different cost rations at once.

Often, Receiver Operator Characteristics (ROCs) are used for this. We construct a ROC curve by computing the convex hull of the foreground/background classification results for a selection of parameter settings, see [19, 20]. Given a certain cost fraction, the optimal method can now easily be found in the graph. We call this curve the total–ROC.

Before foreground/background classification, the EM model was initialized by updating the model for images 500, 499, and so on until image 1 using a constant update speed $u_\mathrm{B} = 0.05$. As the first frame on which we evaluate is frame 500, this allows the algorithm to initialize during 1000 frames. This enables the evaluation of very low update speeds, for which the sequences would not be long enough to obtain a converged model of the background. The entire image sequence was processed several times. Each time using different values for the update speed u_B and threshold T_Stauffer.

In figure 6 total–ROC curves for a number of methods are given. This are the convex hulls of the results of experiments in which both the update speed u_B and the classification threshold T_Stauffer of the background modelling and classification algorithm are varied. As most methods coincide, we selected only a few methods: the reference method No correction, MofQ with and without outlier removal and QofA with outlier removal. For the PETS01-3TR1 and Schiphol sequences the improvement with any of the proposed methods is significant, for the Intratuin sequence there is also a slight improvement.

Table 3 gives an overview of the surface above the ROC for all methods. For Intratuin, most methods that do some kind of outlier removal, i.e. statistical outlier removal or using the median, perform slightly better than without global intensity correction. Only exception is QofM without outlier removal. For Schiphol and PETS01-3TR1 the methods with outlier removal perform significantly better with an error reduction of a factor two and three respectively. Best performance is obtained using MofQ without statistical outlier removal.

5.2.3 Parameter Invariant Performance Evaluation (PIPE)

Unfortunately, the total–ROC is in this case not a sufficient criterion. The problem is that the effect of global intensity differences can be partially solved by faster updating of the background, at the cost of robustness. A method might perform well on one image sequence with a certain update speed, but this does not mean it will perform that well with equal update speed on another set of images, recorded under different conditions. This is a matter of robustness

Table 3: ROC results for real image sequences. Given is the average of the surface above the total–ROC in percentages (lower is better).

Data:	Intratuin		Schiphol		PETS01-3TR1		Average	
Outlier removal:	No	Yes.	No	Yes	No	Yes	No	Yes
Method:								
No	3.95		2.59		2.28		2.94	
QofA	4.19	3.67	2.71	1.38	0.80	0.81	2.57	1.95
QofM	5.38	3.71	1.35	1.34	0.82	0.80	2.51	1.95
AofQ	4.21	3.68	2.03	1.35	0.82	0.79	2.35	1.94
MofQ	3.68	3.68	1.25	1.32	0.76	0.78	1.90	1.93
LS	4.44	3.69	3.01	1.42	0.81	0.80	2.75	1.97
WLS	7.67	3.68	6.08	1.60	1.16	0.80	4.97	2.03

against setting the parameters, specifically the update speed, that is not taken into account by comparing total–ROCs.

We therefore propose the Parameter Invariant Performance Evaluation (PIPE) criterion. PIPE is a measure that gives in one number the average performance on an image sequence and the robustness against changing the parameters to the optimal parameter settings for other image sequences. A low error can be achieved by a method that performs well with all parameter settings, or one that has optimal performance for one setting regardless of the image sequence. Both cases are attractive to use in practice, where it is difficult and impractical to tune parameters as the circumstances change.

PIPE is based on ROCs. The parameters we intend to vary are the threshold T_{Stauffer} and the update speed u_{B}, where the update speed is the parameter we wish to be robust against. We first analyze the effect of this parameter. Therefore, we create for each image sequence l_{seq} and update speed u_{B} a u_{B}–ROC by varying threshold T_{Stauffer} only. We calculate area $A(l_{\text{seq}}, u_{\text{B}})$ under this u_{B}–ROC. This area lies between zero and one, where one corresponds to the perfect classification result and a value of 0.5 can be obtained by classifying at random.

Figure 7 shows the areas $A(l_{\text{seq}}, u_{\text{B}})$ under the u_{B}–ROC for different update speeds for the methods selected before. This immediately shows that different image sequences can require different parameter settings, but that intensity correction makes the algorithm more robust against setting of the update speed. As a consequence, intensity correction makes the algorithm more robust against changes of the environment. The figure also shows that intensity correction cannot be substituted by faster updating of the background model. For the PETS01-3TR1 sequence this might be a solution, but for the Intratuin and Schiphol sequences performance significantly drops for higher update speeds due to misclassification of slowly moving objects.

We wish to obtain one error measure for each image sequence. This can be achieved by averaging the results of the u_{B}–ROC over the different values of the update speed. In order to introduce robustness into the evaluation criterium, we weight the different contributions. The weight is determined by the number of times this update speed was optimal for any of the image sequences. A value of the update speed is optimal when it lies on the convex hull of the total–ROC. $U(u_{\text{B}})$ is the average over all image sequences of the normalized histograms of the occurrence frequency of update speeds on the convex hulls of the total–ROCs. See figure 5.2.3 for some of these histograms.

Our Parameter Invariant Performance Evaluation error e_{PIPE} is now given by one minus this weighted average

$$e_{\text{PIPE}}(l) = 1 - \frac{1}{N_{u_{\text{B}}}} \sum_{u \epsilon n_{u_{\text{B}}}} U(u_{\text{B}}) A(l, u_{\text{B}}) \ , \tag{18}$$

with $n_{u_{\text{B}}}$ the list of $N_{u_{\text{B}}}$ different values of u_{B} that are used.

5.2.4   PIPE results

Full results according to e_{PIPE} described above are given in table 4. Error reductions of almost a factor two on the Schiphol sequence and almost a factor four on the PETS01-3TR1 sequence are obtained. Below we will discuss the results for all sequences in more detail.

For the Intratuin sequence, results of combinations that either do statistical outlier removal or are based on the median are slightly better then the result without correction, except for QofM without statistical outlier removal. It is not surprising that the results are close to that without correction as this image sequence does not contain significant intensity variations. However, this data does show that the use of intensity correction does not make results worse when there is no need for correction. Methods that perform best are AofQ and QofA with statistical outlier removal and MofQ.

The Schiphol sequence contains considerable changes in intensity due to automatic gain control. The error is reduced by approximately a factor one-and-a-half for methods that do some kind of outlier removal. Without outlier removal results are significantly worse. Best performance is obtained using MofQ.

The PETS01-3TR1 sequence contains changes in intensity caused by moving clouds. Therefore, these changes are not always global. Nevertheless, results improve significantly when using any of the the proposed global intensity correction algorithms. An error reduction of almost a factor four is obtained, the methods based on the median perform best on this image sequence. Outlier removal does not seem necessary for the LS method on this image sequence. This is because compared to the other data, there are much less outliers in this image sequence as only a small fraction of the pixels depict foreground. In this case outlier removal still reduces the number of pixels and consequently the estimation accuracy of the apparent gain factor.

On average, outlier removal is essential for good results. MofQ performs best with an average error reduction of more than a factor 1.6.

5.3 Runtime

Our goal is to find an intensity correction algorithm to be used as pre-processing for real-time applications. Therefore, runtime is an important issue.

We ran all algorithms on an image of 100,000 pixels. The runtime of each algorithm was averaged over 100 runs. The computations were performed on a Pentium IV 2500 MHz processor under Windows 2000 Professional. The algorithms were implemented in Matlab 6.1 (www.mathworks.com). The Matlab implementations for the computation of the mean and median were used and all computations were performed in double precision.

The results are given in table 5. This table shows that when no statistical outlier removal is necessary, QofA and AofQ are most affordable, followed by LS. These methods are a good choice in conjunction with other ways of outlier removal like foreground/background classification.

Otherwise statistical outlier removal should be used, except for methods based on the median as we have shown. We therefore compare the methods with a median without outlier removal to the other methods with statistical outlier removal. Then MofQ is the most affordable algorithm, followed by AofQ and QofA, all very close to each other.

It should be noted that it is not necessary to use all pixels in the image for intensity estimation. According to the amount of processing power available and the accuracy requested, a fraction of the available pixels can be used. Outlier removal also reduces the number of pixels that needs to be taken into account.

Table 4: PIPE results for real image sequences. Given is the average of the error e_{PIPE} in percentages (lower is better).

Data:	Intratuin		Schiphol		PETS01-3TR1		Average	
Outlier removal:	No	Yes.	No	Yes	No	Yes	No	Yes
Method:								
No	8.66		6.15		5.04		6.61	
QofA	7.92	7.39	5.13	3.97	1.53	1.46	4.86	4.27
QofM	8.96	7.54	3.83	3.90	1.39	1.41	4.73	4.28
AofQ	7.94	7.39	4.92	3.83	1.48	1.47	4.78	4.23
MofQ	7.14	7.23	3.47	3.63	1.41	1.35	4.01	4.07
LS	8.50	7.67	5.38	4.16	1.45	1.45	5.11	4.43
WLS	12.32	7.90	10.17	4.44	1.76	1.49	8.08	4.61

Table 5: The runtime of the different methods. Times are in milliseconds, standard deviation is lower than 1.0 ms for all measurements.

method:	WLS	LS	QofA	QofM	AofQ	MofQ
No outlier removoval	36.1	10.2	2.5	52.0	6.1	30.6
Statistical outlier removoval	73.1	46.9	39.4	89.1	38.3	62.7

5.4   Discussion of the experiments

According to simulation results, many algorithms seem to perform very good. Results on real images shall therefore be used to select between algorithms. We will look at the average ROC and PIPE performance over all image sequences. We then see that MofQ performs best. This is also the least computational complex algorithm. AofQ and QofA with statistical outlier removal are slightly more expensive, and their performance is also lower.

The use of WLS or LS is expensive and these methods are sensitive to remaining outliers. They are therefore not a good choice if perfect outlier removal cannot be guaranteed. Even though their average performance is good.

Both AofQ or MofQ use a per-pixel estimate of the gain factor. This allows for easy extension to local intensity correction. On the other hand, QofA and QofM have the advantage that they can also be used when the images cannot be compared pixel-wise, like in stereo vision, or when there is a small movement of the camera between the current image and the reference image.

In practice, the application requiring intensity correction should be considered before choosing a method. Considerations are the required precision, the available computing power, whether or not the camera is static and whether or not there is an outlier removal algorithm available. If images can be compared pixel-wise, MofQ is the best choice.

6   Conclusions

6.1   Comparison to previous work

Using simulation we have shown that the proposed intensity correction algorithms perform equally well as can be obtained using ground truth correction. Thus, under the assumptions of constant gamma and only global changes in intensity, there is no need to use more expensive algorithms like dynamic histogram warping [2], iterative weighted least squares estimation of the bias, gain and gamma [25] and estimation of the gain and bias together with optic flow estimation [1]. Also, [2] and [1] do not perform outlier removal. We have shown in this paper that not using outlier removal significantly reduces accuracy and usability.

Our experiments on real data clearly show that only relying on the adaptation of the background modelling technique [15, 24] is not an option. Tweaking the update speed to reach acceptable classification performance on one sequence immediately degrades the performance on other image sequences.

[6] and [14] estimate the intensity change between two images using the average and a least squares estimate respectively. However, these methods do not consider the effect of outliers, reducing accuracy and usability.

6.2   Experimental evaluation

For the simulated images the RMS error between the corrected image and the ground truth image evaluates the accuracy. Results show that methods that use some kind of outlier removal (either statistical or using the median) performed very close to the noise level for image sequences

without and with changes in intensity. This indicates that these methods are sufficiently accurate to be used in practical applications.

Usability of the different methods is demonstrated using background classification on the corrected image sequences. With constant image intensity, all methods with outlier removal obtained equal performance compared to using no correction. For image sequences with changing intensity these methods perform a factor ten better compared to no correction. This is similar to correcting with ground truth values.

ROC and PETS results on real images without temporal intensity changes show that most methods performing outlier removal have slightly better classification performance as without intensity correction. Two image sequences with intensity changes show that the proposed methods cause significant improvement to the classification performance. Even with local intensity changes, significant improvements are shown. Error reduction of almost a factor two on the Schiphol sequence and more than a factor three on the PETS01-3TR1 sequence are obtained.

Using the proposed Parameter Invariant Performance Evaluation PIPE we have further shown that our proposed intensity correction introduces robustness against varying the update speed of the adaptive background modelling algorithm. It allows lower update speed to be used, preventing slowly moving objects to be incorporated in the background model.

6.3   General conclusions

In this paper is shown that using global intensity correction based on the ratio of pixels in conjunction with outlier removal significantly improves background classification results. For image sequences not needing intensity correction, no decrease of performance is seen.

For the evaluation on real images the Parameter Invariant Performance Evaluation (PIPE) error measure, based on the ROC, is proposed. It combines both classification error and robustness against changes in parameter settings in one number.

It is believed that the proposed method can be applied and will provide similar benefit to other image processing algorithms based on Constant Image Brightness (CIB) or the Brightness Constancy Constraint Equation (BCCE).

A   Theoretical model of a CCD camera

Healey [7] describes the following model for a single pixel recorded at time t using a CCD camera:

$$i_t = g_t(h_t i_0 + \mu_{\mathrm{DC}} + N_{\mathrm{S}} + N_{\mathrm{R}}) + N_{\mathrm{Q}} , \tag{19}$$

with $h_t i_0$ the measured scene intensity and i_t the image intensity. The following noise contributions are present[2]: the dark current μ_{DC} is an offset, constant over time. The shot noise N_{S} has a Poisson distribution with $\mu_{\mathrm{S}} = 0$ and σ_{S} depending on i_0. The readout noise N_{R} has a Gaussian distribution: $\mu_{\mathrm{R}} = 0$, σ_{R} constant. The quantization noise N_{Q} has a uniform distribution $U(-\frac{q}{2}, \frac{q}{2})$ with q the smallest step in pixel value.

There are three ways to control the global image intensity, we will use the term apparent gain for their joint effect. It can be controlled using the camera gain g_t, or using camera shutter time or lens iris, modelled together using h_t. All can be fixed (manual control) or automatically adapted to the scene (automatic gain/shutter/iris control).

Additionally, most cameras apply a gamma adjustment to map the range of intensity values from the CCD to the available output range. Assuming it is implemented in the camera electronics just before digitization, equation 19 changes to

$$i_t = g_t^{\gamma} \left(h_t i_0 + \mu_{\mathrm{DC}} + N_{\mathrm{S}} + N_{\mathrm{R}} \right)^{\gamma} + N_{\mathrm{Q}} , \tag{20}$$

---

[2]We use μ_x for the mean of x and σ_x for its standard deviation.

with γ the gamma value which is assumed to be time constant and equal for all pixels.

A.1   Experimental validation of the CCD model

In this section we describe experimental evaluation of the model introduced in equation 20. Besides verification of the model we intend to show the effect of the individual contributions in the model to allow for the simplifications given in section A.2. In particular, we need to know:

- whether the gamma function used in the camera is sufficiently accurately described by our model,

- what the effect of the dark current is,

- what the noise distribution is.

First, we will discuss the measurement setup. Then the gamma, dark current and noise parameters are estimated. Finally, some artifacts and additional measurements are discussed.

A.1.1   Measurement setup

In order to answer the questions given above we did measurements with a range of cameras. A list of the cameras used is given in table 6. As measurement object we used a half-transparent plate which is homogeneously back-illuminated. In front of this plate are layers of gray and brown filters with different thickness, see figure 8. There are nineteen different sections, zero to five layers of gray filter on the top row and zero to thirteen layers of brown filter on the center and bottom row. The intensity of each of the nineteen different sections differs, its average intensity is measured by a photon counter.

Imagery depicting this object was recorded with several cameras. Different sequences were recorded, each sequence consisting of 100 images to allow reducing the noise by averaging. For different sequences, different parts of the object were covered. This changes the covered part of the scene to black, while leaving the remainder of the scene unchanged. This will trigger the automatic gain control of the cameras used while leaving some sections to measure on.

A.1.2   Gamma estimation

For each sequence of 100 images the standard deviation and average of the red color channel were calculated per pixel and of the sections with equal intensity. The gamma can now be estimated from a log-log plot of the average intensity per section against the true (photon counter measured) intensity of that section. All data points should lie on a straight line (at least for points with intensity much greater than the dark current) with a slope equal to gamma. An example of such a graph is depicted in figure 9 top left. The estimated gammas for all cameras are given in table 7. As expected, all cameras have a gamma smaller than one and the individual data points are close to a straight line. Therefore we conclude that *our model of the gamma is sufficient.*

A.1.3   Dark current estimation

Using the gamma estimated above we correct the image sequences and recalculate the average intensity for each section. Plotting these against the true intensity of the sections should now give a straight line which intersects the section-intensity axes at a value related to the dark current, see figure 9 bottom left. The estimated dark current and the error of the estimate are given in table 7. The dark current we estimated is for all cameras smaller then the standard deviation of the estimate, so we conclude that for pixels with a sufficiently large intensity *we can neglect the dark current.*

Table 6: This table lists the cameras used to verify the CCD model.

Camera	Description	Gain control
JAI CV-M7+CL	High end 10-bit digital camera with Bayer filter	Fixed
Siemens C810	Digital computer vision color camera	Automatic
JVC GR-DVL-157	Digital consumer color video camera	Automatic
Philips PCVC680K	Color webcam	Automatic

Table 7: Results of the camera characterization for each of the cameras. All estimates are least squares estimates of the available data. Between brackets are the standard deviations of the estimates. The amount of multiplicative noise is given for the highest pixel value (one in our case).

Camera	JAI	Siemens	JVC	Philips
Gamma	0.66 (0.02)	0.53 (0.02)	0.72 (0.06)	0.70 (0.10)
Dark Current $g_t h_t \mu_{DC}$	-0.003 (0.005)	-0.0005 (0.009)	0.005 (0.016)	0.005 (0.03)
Add. noise $g_t N_R$	0.003	0.003	0.005	0.012
Multipl. noise $g_t N_S$	0.021	0.021	0.019	0.011
Quant. noise N_Q	0.0003	0.001	0.001	0.001

A.1.4 Noise estimation

The distribution of the noise in the CCD model contains both additive and multiplicative noise. We plot the standard deviation over the gamma-corrected images against its average for a number of pixels to see the effect of both contributions, see figure 9 top right. The intersection of a straight line fitted to this data with the standard deviation axes gives us the contribution of the additive noise and the slope of the line gives the amount of multiplicative noise. From estimates of the amount of additive and multiplicative noise given in table 7 we conclude that for all cameras except the Philips web cam and for pixels with a sufficiently large intensity *we can neglect the additive part of the noise.*

A.1.5 Artifacts and additional measurements

The Philips webcam seems to have some artifacts (among others a gamma that changes as the intensity changes, violating the general model given in equation 20). Therefore our proposed algorithms might not work correctly on data recorded with such a camera.

Experiments were repeated with different intensity-filters directly in front of the camera lens to evaluate the effect of lower overall intensity. These additional experiments show the same effects and therefore also support the conclusions given above. However, their results are not given here.

In subsection 5.1 simulated images are described. These are based on the parameters as we measured them in this appendix:

- $\gamma = 0.65$

- $\mu_{DC} = 0.07$ (worst case from the measured cameras)

- $\sigma_S = 0.02$

- $\sigma_R = 0.005$.

A.2   Simplifications to the model

The previous section reports the experimental results of the different contributions in this CCD model. These results allow for simplification of the model. The simplifications are given in this subsection and lead to the model of the CCD camera to be used in the remainder of this paper.

Our experimental validation concludes that we can neglect the contribution of the dark current. This simplifies equation 20 to

$$i_t = g_t^\gamma \left(h_t i_0 + N_\mathrm{S} + N_\mathrm{R}\right)^\gamma + N_\mathrm{Q} \ . \tag{21}$$

The remaining noise terms all are zero-mean. The shot noise N_S is multiplicative and the readout noise N_R and the quantization noise N_Q are additive.

Our experiments also showed that the additive noise contributions (N_R and N_Q) are negligible compared to the multiplicative noise contribution for sufficiently large intensity values. This simplifies the above equation to

$$i_t = g_t^\gamma \left(h_t i_0 + N_\mathrm{S}\right)^\gamma \ . \tag{22}$$

References

[1] Yucel Altunbasak, Russel M. Mersereau, and Andrew J. Patti. A fast parametric motion estimation algorithm with illumination and lens distortion correction. *IEEE Trans. Image Process*, 12(4):395–408, 2003.

[2] Ingemar J. Cox, Sibastien Roy, and Sunita L. Hingorani. Dynamic histogram warping of image pairs for constant image brightness. In *Proc. IEEE Int'l Conf. Image Processing (ICIP)*, pages 2366–2369, 1995.

[3] Michael Greiffenhagen, Visvanathan Ramesh, and Heinrich Niemann. The systematic design and analysis cycle of a vision system: A case study in video surveillance. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages II:704–711, 2001.

[4] Gregory D. Hager and Peter N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans. Pattern Anal. Machine Intell.*, 20(10):1125–1139, 1998.

[5] Michael Harville, Gaile G. Gordon, and John Woodfill. Foreground segmentation using adaptive mixture models in color and depth. In *Proc. IEEE Workshop Detection and Recognition of Events in Video (WDREV)*, pages 3–11, 2001.

[6] Yinghua He, Hong Wang, and Bo Zhang. Background updating in illumination–variant scenes. In *Proc. IEEE Int'l Conf. Intelligent Transportation Systems (ITSC)*, pages I:515–519, 2003.

[7] Glenn E. Healey and Raghava Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Trans. Image Process*, 16(3):267–276, 1994.

[8] Horst Hischof, Horst Wildenauer, and Ales Leonardis. Illumination insensitive recognition using eigenspaces. *Comput. Vis. Image Underst.*, 95(1):86–104, 2004.

[9] Thanarat Horprasert, David Harwood, and Larry S. Davis. A statistical approach for real–time robust background subtraction and shadow detection. In *Proc. IEEE ICCV'99 FRAME-RATE Workshop*, pages II:751–767, 1999.

[10] Jun-Wei Hsieh, Chia-Jung Chang, Wen-Fong Hu, and Yung-Sheng Chen. Shadow elimination for effective moving object detection by Gaussian shadow modeling. *Image and Vision Computing*, 21(6):505–516, 2003.

[11] Sumer Jabri, Zoran Duric, Harry Wechsler, and Azriel Rosenfeld. Detection and location of people in video images using adaptive fusion of color and edge information. In *Proc. IEEE Int'l Conf. Pattern recognition (ICPR)*, pages IV:627–630, 2000.

[12] David W. Jacobs, Peter N. Belhumeur, and Ronen Basri. Comparing images under variable illumination. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 610–617, 1998.

[13] Omar Javed and Mubarak Shah. Tracking and object classification for automated surveillance. In *Proc. European Conf. Computer Vision (ECCV)*, pages IV:343–357, 2002.

[14] K. Kamikura, Hiroki Watanabe, H. Jozawa, H. Kotera, and S. Ichinose. Global brightness–variation compensation for video coding. *IEEE Trans. Circuits and Systems*, 8(8):988–1000, 1998.

[15] Stephen J. McKenna, Sumer Jabri, Zoran Duric, Azriel Rosenfeld, and Harry Wechsler. Tracking groups of people. *Comput. Vis. Image Underst.*, 80(1):42–56, 2000.

[16] Naoya Ohta. A statistical approach to background subtraction for surveillance systems. In *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, pages II:751–767, 2001.

[17] Ioannis Pavlidis and Vassilios Morellas. Two examples of indoor and outdoor surveillance systems: motivation, design, and testing. In *Proc. European Workshop Advanced Video Based Surveillance Systems (AVBS)*, pages 285–296, 2001.

[18] Carey E. Priebe. Adaptive mixtures. *J. of the American Statistical Association*, 89(427):796–806, 1994.

[19] Foster J. Provost and Tom Fawcett. Robust classification for imprecise environments. *Machine Learning*, 42(3):203–231, 2001.

[20] M.J.J. Scott, M. Niranjan, and Richard W. Prager. Realisable classifiers: Improving operating performance on variable cost problems. In *Proc. Brittisch Machine Vision Conf. (BMVC)*, pages 306–315, 1998.

[21] Andreas Siebert. Retrieval of gamma corrected images. *Pattern Recogn. Letters*, 22(2):249–256, 2001.

[22] Chris Stauffer and W. Eric L. Grimson. Learning patterns of activity using real–time tracking. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(8):747–757, 2000.

[23] D. Toth, Ti Aach, and V. Metzler. Illumination–invariant change detection. In *Proc. IEEE Southwest Symposium Image Analysis and Interpretation (SWSYMP)*, pages 3–7, 2000.

[24] Kentaro Toyama, John Krumm, Barry Brumitt, and Brian Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, pages 255–261, 1999.

[25] Yanghai Tsin, Visvanathan Ramesh, and Takeo Kanade. Statistical calibration of CCD imaging process. In *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, pages I:480–487, 2001.

[26] Paul J. Withagen, Frans C.A. Groen, and Klamer Schutte. CCD characterization for a range of color cameras. In *Proc. IEEE Instrumentation and Measurement Technical Conference (IMTC)*, pages III:2232–2235, 2005.

[27] Paul J. Withagen, Frans C.A. Groen, and Klamer Schutte. Shadow detection using a physical basis. Technical Report IAS-UVA-07-02, Intelligent Autonomous Systems, University of Amsterdam, The Netherlands, January 2007.

[28] Paul J. Withagen, Klamer Schutte, and Frans C.A. Groen. Probabilistic classification between foreground objects and background. In *Proc. IEEE Int'l Conf. Pattern recognition (ICPR)*, pages 31–34, 2004.

[29] Binglong Xie, Visvanathan Ramesh, and Terrance E. Boult. Sudden illumination change detection using order consistency. *Image and Vision Computing*, 22(2):117–125, 2004.
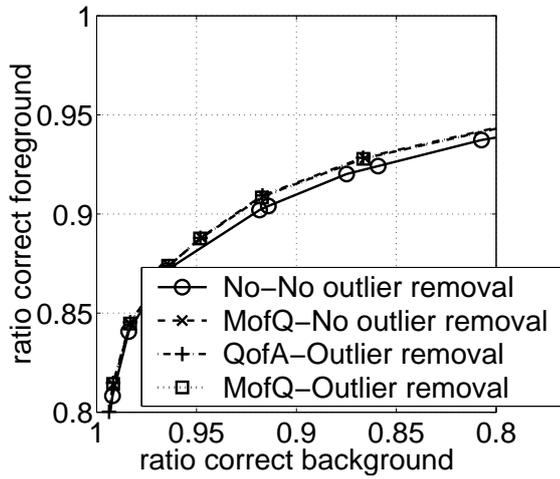
(a) Intratuin
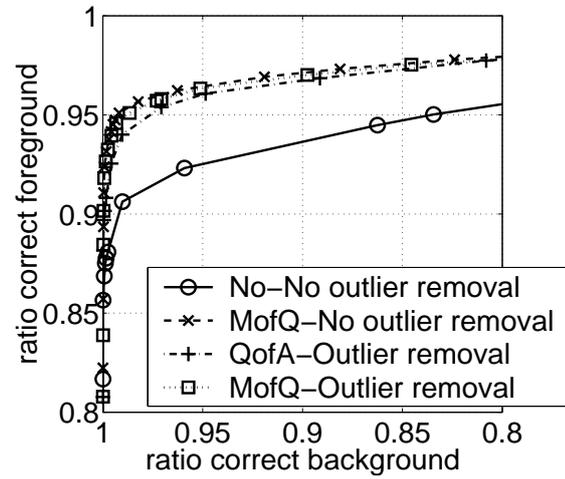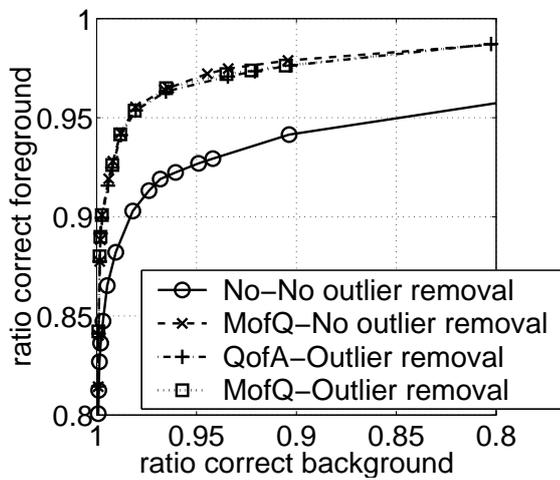
(b) Schiphol

(c) PETS01-3TR1

Figure 6: Total–ROCs of the classification results on real data with different values of update speed and threshold. Only a selection of the methods is shown as most methods coincide.
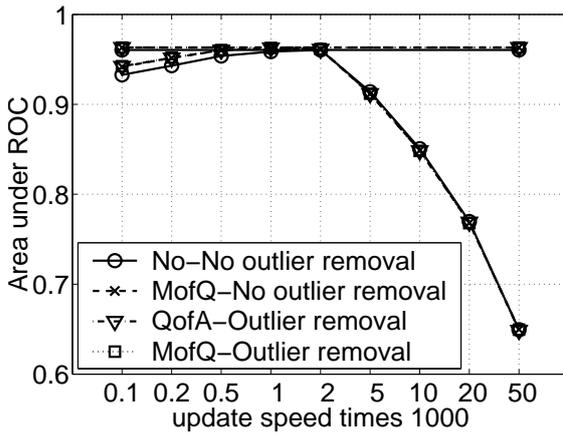
(a) Intratuin
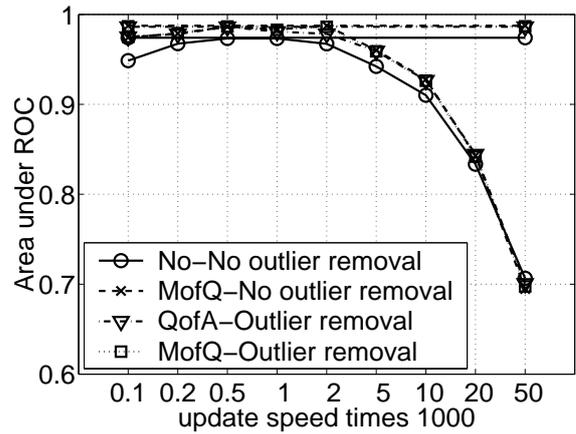
(b) Schiphol
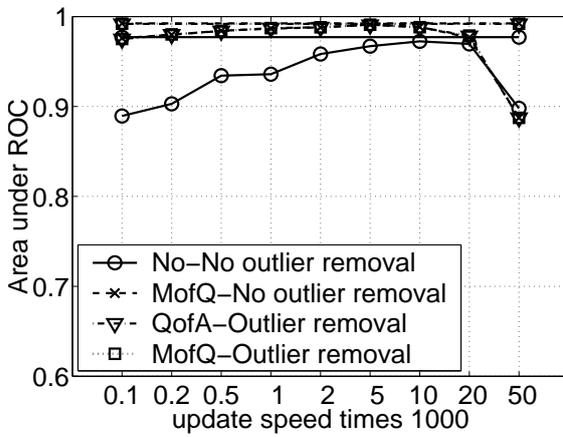
(c) PETS01-3TR1
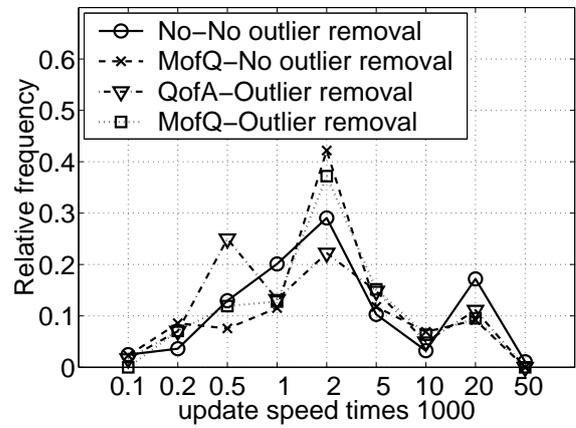
(d) $U(u)$ histogram

Figure 7: The area under the u_{B}–ROC for different values of the update speed. The horizontal lines indicates the area under the total–ROC. Only a selection of the methods is shown as most methods coincide. Figure (d) shows the corresponding $U(u_{\mathrm{B}})$ histograms.

(a) Back-illumination off (JVC)
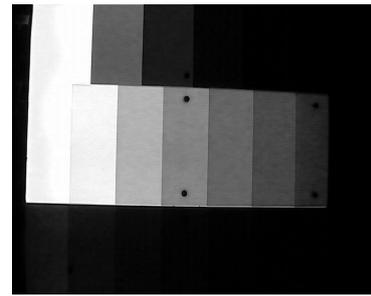
(b) Back-illumination on, red channel (JVC)

Figure 8: Two pictures of the measurement object.
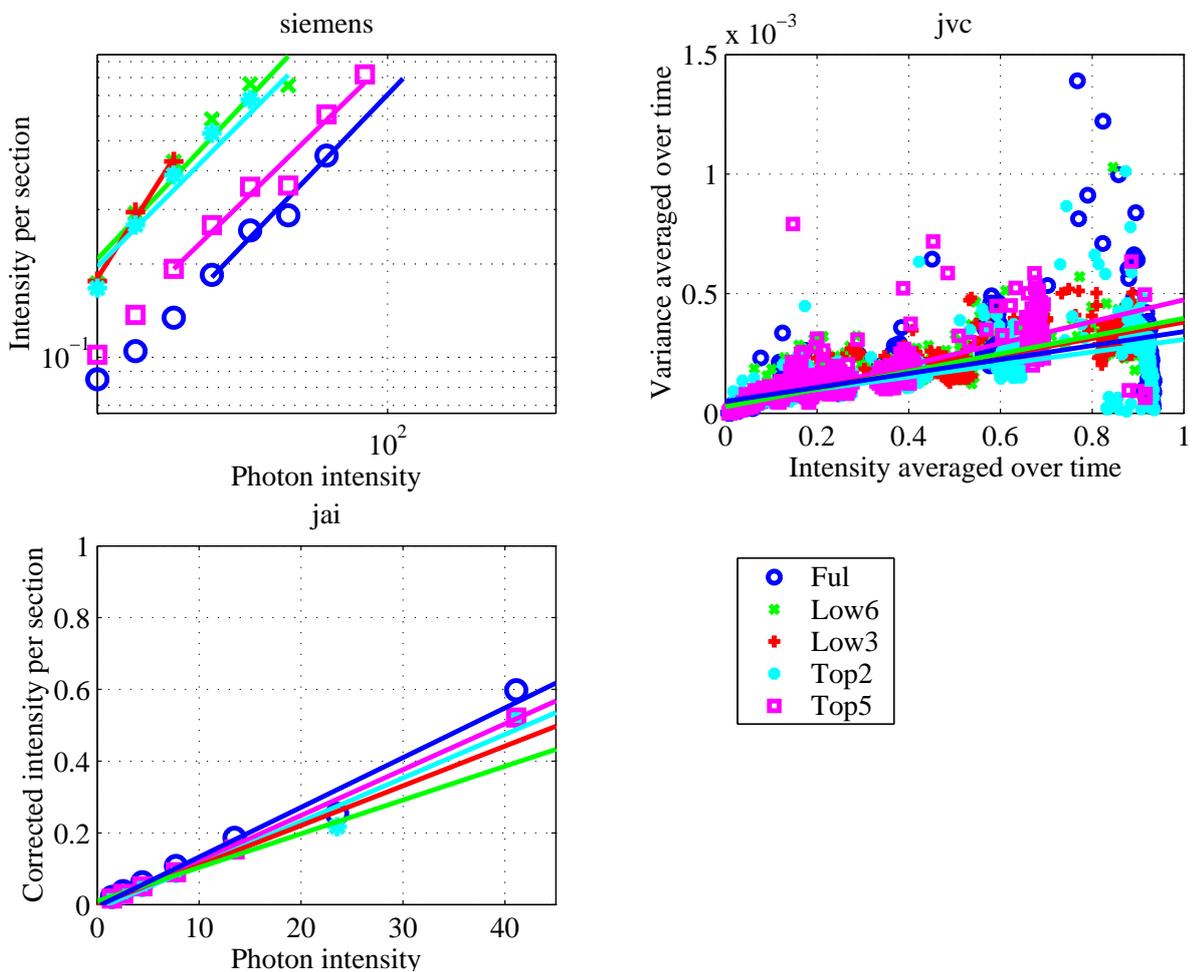
Figure 9: An example of each of the graphs mentioned. Each graph is shown for a different camera. On the top left we see a log-log plot of the image intensity per section against the true intensity of the sections. The bottom left graph shows a plot of gamma-corrected image intensity per section against the true section intensities. The top right graph shows for a number of pixels the standard deviation over time plot against the average over time. See [26] for more figures.

Acknowledgements

While this research was conducted, PW was affiliated with [2] and [3]

IAS reports

This report is in the series of IAS technical reports. The series editor is Bas Terwijn (`bterwijn@science.uva.nl`). Within this series the following titles appeared:

A. Noulas and B. Kröse *Probabilistic audio visual sensor fusion for speaker detection.* Technical Report IAS-UVA-06-04, Informatics Institute, University of Amsterdam, The Netherlands, March 2006.

J. Nunnink and G. Pavlin *Fault localization in bayesian networks.* Technical Report IAS-UVA-06-03, Informatics Institute, University of Amsterdam, The Netherlands, May 2006.

F. Oliehoek and N. Vlassis *Dec-POMDPs and extensive form games: equivalence of models and algorithms.* Technical Report IAS-UVA-06-02, Informatics Institute, University of Amsterdam, The Netherlands, April 2006.

All IAS technical reports are available for download at the IAS website, `http://www.science.uva.nl/research/ias/publications/reports/`.