Global illumination compensation for background subtraction using Gaussian-based background difference modeling

Citation for published version (APA):

Vijverberg, J. A., Loomans, M. J. H., Koeleman, C. J., & With, de, P. H. N. (2009). Global illumination compensation for background subtraction using Gaussian-based background difference modeling. In *6th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2009, 2 - 4 September 2009, Genova* (pp. 448-453). Institute of Electrical and Electronics Engineers.
<https://doi.org/10.1109/AVSS.2009.101>

DOI:

[10.1109/AVSS.2009.101](https://doi.org/10.1109/AVSS.2009.101)

Document status and date:

Published: 01/01/2009

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Global Illumination Compensation for Background Subtraction Using Gaussian-based Background Difference Modeling

Julien A. Vijverberg^{1,2}, Marijn J.H. Loomans^{1,2}, Cornelis J. Koeleman¹, Peter H.N. de With^{2,3}

{julien,marijn,rick}@vdg-security.com, p.h.n.de.with@tue.nl

¹VDG Security B.V., Zoetermeer, The Netherlands

²Eindhoven University of Technology, Eindhoven, The Netherlands

³CycloMedia Technology B.V., Waardenburg, The Netherlands

Abstract

This paper presents a background segmentation technique, which is able to process acceptable segmentation masks under fast global illumination changes. The histogram of the frame-based background difference is modeled with multiple kernels. The model that represents the histogram at best, is used to determine the shift in luminance due to global illumination or diaphragm changes, such that the background difference can be compensated. Experimental results have revealed that the number of incorrectly classified pixels using global illumination compensation instead of only the approximated median method reduces from 77% to 19% shortly after a fast change. The performance of the proposed technique is similar to state-of-the-art related work for global illumination changes, despite the fact that only luminance information is used. The algorithm is computationally simple and can operate at 30 frames-per-second for VGA resolution on a P-IV 3-GHz PC.

1. Introduction

The continuing trend to increase the computation power of processors, allows a new generation of security cameras to be equipped with powerful Digital Signal Processors (DSPs) and/or reconfigurable logic like Field Programmable Gate Arrays (FPGAs). These DSPs and FPGAs enable security cameras to improve the compression of high-resolution video and allows content analysis algorithms to detect relevant events.

In video-content analysis such as applied within surveillance systems, background segmentation is commonly used to extract the relevant objects from an input image. Important requirements for background segmentation include that (1) the number of pixels correctly classified should be high,

(2) the background should be updated regularly, while relevant objects should not distort the background and (3) the computational complexity should be low. Examples where traditional background segmentation techniques tend to fail are background objects that are changed or removed, vegetation moved by the wind, variations in weather, illumination changes and stationary objects.

Since stationary and slowly moving objects tend to appear in the background image, the quality of the segmentation mask degrades. A simple counter measure is to decrease the update rate of the background estimation algorithm. However, when the background changes quickly (often due to global illumination changes), the slowly-updated background image loses its validity for the current situation. Therefore, we need to model and compensate the global illumination changes in the sequence.

In literature on background segmentation, Stauffer and Grimson [13] and Elgammal *et al.* [3] were the first in using advanced density estimation techniques on pixel-level; Gaussian Mixture Model (GMM) and non-parametric techniques respectively. No global difference models were used. In [14], Toyama *et al.* present the Wallflower background estimation, which estimates multiple backgrounds (two in their experiments) and switches between the learned backgrounds when a large number of pixels are detected as foreground, for example due to different illumination conditions. This is a very pragmatic approach, but it does only provide a solution for two states, which also have to be trained in advance.

This has led others [5, 7] to follow a more structural approach. In these papers, gradient information was used to classify foreground pixels in two categories: ‘illumination change’ and ‘object present’. Gradient information is most discriminating at the boundaries of the objects, but does not provide a clear difference for large, untextured objects (*e.g.* a white bus on the road surface). Furthermore, gradient in-

this reason, we have decided to fit these three models on the histogram and select the best matching model.

In the sequel, a description is given of how the three models are fit onto the histogram h_Δ (Section 3.1), the selection of the best model (Section 3.2) and how the background difference is compensated to obtain a segmentation mask (Section 3.3).

3.1. Illumination Parameter Estimation

First, the parameters $\theta = \{w, \mu, \sigma\}$ for a Laplacian distribution

$$\hat{h}_\Delta[u; \theta] = we^{-|u-\mu|/\sigma}, \quad (1)$$

and a Gaussian-shaped function

$$\hat{h}_\Delta[u; \theta] = we^{-(u-\mu)^2/2\sigma^2}, \quad (2)$$

are computed. For the both histogram models, the μ parameter is set to the median of the measured histogram. For the Laplacian distribution, σ is set to $\sum_u |h_\Delta[u] - \mu|/N_s$ and w is set to $\frac{1}{2\sigma} \sum_u h_\Delta[u]$. For the Gaussian distribution, we have initialized σ with the square root of the variance, and w with $\frac{1}{\sqrt{2\pi}\sigma} \sum_u h_\Delta[u]$.

Second, we estimate the parameters of a two-component GMM. A two-component GMM better covers the diaphragm changes and complicated difference histograms. We apply the Levenberg-Marquardt (LM) algorithm to estimate the parameters and use the implementation from [8]. This optimization algorithm iteratively finds the minimum Sum of Squared Distances (SSD) using the Jacobian of the SSD. For our experiments, the parameters of the components are initialized, using the parameters of the single Gaussian function $\theta_{1,2} = \{\frac{w}{2}, \mu \pm 10, \sigma/\sqrt{2}\}$.

3.2. Illumination Model Selection

Once the parameters of the models are estimated, the SSD is computed for each of them. As the SSD is a type of error metric, it is directly clear that the one-component model with the largest SSD should not be selected. However, since the two-component model is likely to better fit the histogram h_Δ , we need a method to select either the one-component or two-component model, while penalizing the two-component model for the additional parameters. Both the Bayesian Information Criterion (BIC) and Akaike's Information Criterion (AIC) provide a method to penalize additional parameters [4]. Because experiments showed no large differences in performance using the BIC or AIC, for our experiments we have adopted the AIC specified by

$$\text{AIC} = N_s \log_e (\text{SSD}/N_s) + N_p, \quad (3)$$

where $N_s = 511$ is the number of bins in the histogram and $N_p \in \{3, 6\}$ the number of parameters of the model. According to [4] many authors applied the AIC to non-linear

Figure 2. The number of iterations of the optimization procedure versus the number of correctly classified pixels in the 'Wallflower' light switch sequence.

problems. For this reason, we have also applied Eq. (3) to our non-linear optimization problem.

3.3. Illumination Compensation

The illumination compensation is performed by computing the thresholds on the background difference to select between foreground (255) and background (0). Suppose that a model was selected with K mixture components and that model k (for $k = 0, \dots, K-1$) has mean μ_k .

$$F(x, y) = \begin{cases} 255 & \text{if } |\Delta(x, y) - \mu_k| > T_k \forall k \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The threshold T_k is set to $\max(T_{\min}, 1.5\sigma_k)$, such that the allowed difference increases with the variance but cannot be smaller than $T_{\min} = 16$. Despite the fact that the additional thresholds increase the complexity of the subtraction step compared to the basic algorithm $|I(x, y) - B(x, y)|$, the overall complexity is still low.

4. Experiments

The maximum number of iterations of the LM optimization is set to 10. Fig. 2 shows that after 10 iterations, the number of correctly classified pixels is almost constant. The peak at 2 and 3 iterations can be explained by the fact that the desk area in the the bottom-right part of the image (see Fig. 4(a)) does not fit the illumination model, but is correctly classified as background during the iterations 2 and 3. In other sequences, we also observed no significant differences between 10 or 100 iterations, but we could not quantify this, due to a lack of ground-truth data

The output of the BGSUB block (Fig. 1) is shown in Fig. 3(a), 4(a) and 5(a) for the 'AVSS PV Medium', the 'Wallflower' light switch and a custom sequence, respectively. Each of these figures shows, from left to right, the input frame, the segmentation mask using illumination compensation and the segmentation mask without illumination compensation.

(a) Input image (left). Mask produced by segmentation with (middle figure) and without (right figure) illumination compensation

(b) MMM updated each frame

(c) MMM updated every 16 frames

(d) MMM updated every 64 frames

Figure 3. Results of the global illumination compensation method for the AVSS PV Medium sequence.

(a) Input image (left). Mask produced by segmentation with (middle figure) and without (right figure) illumination compensation

(b) MMM updated each frame

(c) MMM updated every 16 frames

(d) MMM updated every 64 frames

Figure 4. Results of the global illumination compensation method for the Wallflower light-switch sequence up sampled to 640×480 .

In addition, we compare the results to the segmentation mask generated by Multi-Modal Mean (MMM) [1] which is a state-of-the-art background estimation technique intended for embedded vision systems. In the experiments, we use 4 modes and vary the update period between 1 frame, 16 frames and 64 frames. In the remainder of this paper, these will be abbreviated by MMM-1, MMM-16 and MMM-64, respectively. Figures 2(b-d), 3(b-d), 4(b-d) show results obtained with the MMM technique.

4.1. Accuracy

Fig. 3 shows a still of the ‘AVSS PV Medium’ sequence [6]. This video sequence was recorded with a vibrating camera and contains multiple stationary objects during global illumination changes. Fig. 3(b)-3(d) show that under these circumstances, both MMM and the proposed technique provide segmentation masks of acceptable quality.

Fig. 4 shows the results on the light-switch sequence, which is one of the ‘Wallflower’ sequences [14]. These

(a) Input image (left). Mask produced by segmentation with (middle figure) and without (right figure) illumination compensation

(b) MMM updated each frame (c) MMM updated each 16 frames (d) MMM updated each 64 frames

Figure 5. Results of the global illumination compensation method for a custom sequence.

results show that the proposed technique is better than the uncompensated difference, but even the segmentation mask created using compensation is not very good. However, our results are similar to the results by Pilet *et al.* [11]. This is partly due to the flickering computer screen and the chair being pulled away such that new background is revealed. To compare the results with the ground truth mask, the results are down-sampled to 160×120 pixels. Table 1 shows the number of incorrectly classified pixels.

It has to be noted that the use of a single ground-truth frame has a large impact on the segmentation results. For all update periods, the MMM algorithm adapts the background after the light switch. However, during the frames before the ground-truth frame, MMM-16 adapts slightly faster than the other two. In addition, after MMM has adapted to the new situation (lights turned on/off), it handles the flickering computer screen much better than the AM method and the proposed technique.

Since only some of the sequences available exhibit global illumination changes, we have also captured custom sequences to test the proposed algorithm. Fig. 5 shows the results for our custom sequence, which contains many global illumination changes due to weather conditions and diaphragm changes. Fig. 5(b)-5(d) shows the corresponding segmentation results of the MMM method.

For test sequences not shown in this paper exhibiting less illumination changes, the proposed technique did not noticeably degrade the quality of the segmentation mask. This phenomenon was also noticed for the algorithm experimented by Pilet *et al.* [11]. However, the proposed technique fails on occasions (notably on some parts of the

method	percentage incorrect
MMM-1	60%
MMM-16	19%
MMM-64	56%
w/o compensation	77%
with compensation	19%
Pilet <i>et al.</i>	23%*

Table 1. Percentage of incorrectly classified pixels for the Wallflower light-switch sequence at frame 1865. The percentage generated for Pilet *et al.* is an estimation by the authors based on the numbers in the paper by Pilet *et al.* [11].

‘AVSS PV Medium’ sequence), where the model has not been estimated correctly. This usually happens when the algorithm selects the single Gaussian option, when two Gaussians close to each other would be more appropriate.¹

4.2. Computational Complexity

The algorithm is executed on a single core of a P-IV 2.4-GHz quad-core PC. Table 2 shows the average execution time per frame for each sequence. The execution time includes the time for 3DRS and the complete algorithm as depicted in Fig. 1. The column headed by Time(100) contains the average execution time for a run of upto 100 iterations of the LM algorithm and the column headed by Time(10) for a run of upto 10 iterations. The achieved frame rate for

¹Example sequences are available at <http://vca.ele.tue.nl/demos/avss2009/index.html>

Sequence	Width	Height	Time (100)	Time (10)
AVSS PV Medium	752	576	103	40
Light Switch	640	480	43	32
Custom	512	384	86	16

Table 2. Average execution time (ms) to process a single frame. We distinguish between a maximum of 100 and 10 iterations of the LM optimization procedure to show the relation between the number of iterations and the execution time.

10 iterations varies between 25 and 62 fps. The large differences between the execution time for 100 iterations or 10 iterations imply that the execution time is highly dependent on the number of iterations of the LM algorithm. This illustrates the complexity of the LM algorithm and that our technique can be executed in real time, when the number of iterations is limited.

5. Conclusion

We have presented a technique for more robust background segmentation handling global illumination changes. This technique is based on estimating the PDF of the background difference using two Gaussian models and one Laplacian model. The algorithm selects one of these models on a frame basis, using the AIC and the SSD metric.

In all cases, using illumination compensation yields improvements over the segmentation mask produced by the plain segmentation algorithm. In addition, we have compared the proposed segmentation technique against the MMM background estimation technique. This comparison has shown that MMM can provide a good segmentation mask, if the update rate is tuned to the situation. However, using MMM, stationary objects require special attention, as they may be omitted occasionally. The experiments showed that the proposed technique and MMM are comparable in performance. MMM outperforms our technique in cases of flickering screens and moving vegetation, whereas our technique outperforms MMM during fast global illumination changes. Both techniques appeared to be not very sensitive to camera vibration in our experiments, in contrast to techniques using gradient information.

We have also compared the proposed segmentation technique against the global illumination compensation technique proposed by Pilet *et al.* [11] and found comparable performance. Significant differences are the ability to handle shadows (Pilet *et al.* are able to ignore shadows) and the computational complexity: 6 fps for Pilet *et al.* vs. 21 fps for the proposed algorithm on *e.g.* a 2.0-GHz PC. Finally, we add that Pilet *et al.* require the use of gradient information, which we consider disadvantageous, as it is sensitive to camera vibrations. Besides this aspect, our tech-

nique is more attractive for surveillance cameras with embedded video analysis, since our algorithmic complexity is much lower and we favour the robustness on illumination changes.

References

- [1] S. Apewokin, B. Valentine, D. Forsthoefel, L. Wills, S. Wills, and A. Gentile. *Embedded Computer Vision*, chapter 8, pages 163–175. Springer London, 2008.
- [2] G. de Haan, P. Biezen, H. Huijgen, and O. Ojo. True-motion estimation with 3-D recursive search block matching. *IEEE Trans. on Circuits and Systems for Video Technology*, 3(5):368–379, 1993.
- [3] A. M. Elgammal, D. Harwood, and L. S. Davis. Non-parametric Model for Background Subtraction. In *Proc. of the 6th European Conf. on Computer Vision-Part II*, pages 751–767, 2000.
- [4] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2001.
- [5] M. Heikkilä and M. Pietikäinen. A Texture-Based Method for Modeling the Background and Detecting Moving Objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 657–662, 2006.
- [6] i-Lids. i-Lids dataset for AVSS 2007. http://www.elec.qmul.ac.uk/staffinfo/andrea/avss2007_d.html, September 2007.
- [7] O. Javed, K. Shafique, and M. Shah. A hierarchical approach to robust background subtraction using color and gradient information. In *Proc. Workshop on Motion and Video Computing, 2002*, pages 22–27, 2002.
- [8] M. I. A. Lourakis. levmar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++. <http://www.ics.forth.gr/lourakis/levmar/>, July 2004.
- [9] N. J. B. McFarlane and C. P. Schofield. Segmentation and tracking of piglets in images. *Machine Vision and Applications*, 8(3):187–193, 1995.
- [10] P. Noriega and O. Bernier. Real time illumination invariant background subtraction using local kernel histograms. In *Proc. of the British Machine Vision Association*, 2006.
- [11] J. Pilet, C. Strecha, and P. Fua. Making Background Subtraction Robust to Sudden Illumination Changes. In *Proc. European Conf. on Computer Vision*, October 2008.
- [12] M. Seki, T. Wada, H. Fujiwara, and K. Sumi. Background Subtraction based on Cooccurrence of Image Variations. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages II–65 – II–72, June 2003.
- [13] C. Stauffer and W. Grimson. Adaptive Background Mixture Models for Real-Time Tracking. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 246–252, June 1999.
- [14] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Int. Conf. on Computer Vision*, volume 232, 1999.