

Denoising Fluorescence Endoscopy: A Motion-Compensated Temporal Recursive Video Filter with an Optimal Minimum Mean Square Error Parametrization

Thomas Stehle and Jonas Wulff and Alexander Behrens and
Sebastian Gross and Til Aach

Institute of Imaging and Computer Vision
RWTH Aachen University, 52056 Aachen, Germany
tel: +49 241 80 27860, fax: +49 241 80 22200
web: www.lfb.rwth-aachen.de

in: Proceedings IEEE International Symposium on Biomedical Imaging (ISBI). See also $\text{BIB}_{\text{T}_{\text{E}}\text{X}}$ entry below.

$\text{BIB}_{\text{T}_{\text{E}}\text{X}}$:

```
@inproceedings{STE09c,  
author      = {Thomas Stehle and Jonas Wulff and Alexander Behrens and Sebastian Gross and Til Aach},  
title       = {Denoising Fluorescence Endoscopy: A Motion-Compensated Temporal Recursive Video Filter  
              with an Optimal Minimum Mean Square Error Parametrization},  
booktitle   = {Proceedings IEEE International Symposium on Biomedical Imaging (ISBI)},  
publisher   = {IEEE},  
address     = {Boston},  
month       = {June 28 - July 1},  
year        = {2009},  
pages       = {314--317}}
```

© 2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

DENOISING FLUORESCENCE ENDOSCOPY – A MOTION COMPENSATED TEMPORAL RECURSIVE VIDEO FILTER WITH AN OPTIMAL MINIMUM MEAN SQUARE ERROR PARAMETERIZATION

Thomas Stehle, Jonas Wulff, Alexander Behrens, Sebastian Gross and Til Aach

Institute of Imaging & Computer Vision, RWTH Aachen University, D-52056 Aachen, Germany

ABSTRACT

Fluorescence endoscopy is an emerging technique for the detection of bladder cancer. A marker substance is brought into the patient's bladder which accumulates at cancer tissue. If a suitable narrow band light source is used for illumination, a red fluorescence of the marker substance is observable. Because of the low fluorescence photon count and because of the narrow band light source, only a small amount of light is detected by the camera's CCD sensor. This, in turn, leads to strong noise in the recorded video sequence.

To overcome this problem, we apply a temporal recursive filter to the video sequence. The derivation of a filter function is presented, which leads to an optimal filter in the minimum mean square error sense. The algorithm is implemented as plug-in for the real-time capable clinical demonstrator platform *RealTimeFrame* and it is capable to process color videos with a resolution of 768×576 pixels at 50 frames per second.

Index Terms— Noise filtering, optimal filter, endoscopy, fluorescence, bladder, photo dynamic diagnostics

1. INTRODUCTION

Bladder cancer has its highest incidence rate in industrialized countries. By far, the greatest risk factor for this disease is smoking of tobacco products. Other important risk factors are contact with aromatic amines (e. g. through working in the dye industry), bilharziosis, and age. According to the National Institutes of Health (NIH) approximately 69,000 people were diagnosed with the disease and the number of deaths is 14,150 in 2008 in the United States.

Bladder Cancer can be diagnosed and treated during an endoscopic examination (so-called cystoscopy). A cystoscope is brought through the urethra into the bladder, which is filled with isotonic saline solution. The cancerous tissue can then be removed using endoscopic tools, e. g. a resectoscope cutting loop.

As bladder cancer is very difficult to recognize, fluorescence endoscopy (also called photo dynamic diagnostics - PDD) is used to enhance its visibility. To this end, a marker

substance, like 5-aminolaevulinic acid (5-ALA), is instilled into the patient's bladder two and a half hours before the actual intervention starts. It accumulates within tissues exhibiting high metabolic rates such as tumors. Being exposed to a special blue narrow band illumination, 5-ALA starts fluorescing in red. Therefore, healthy and cancerous tissues can be distinguished more easily (see Fig. 1).

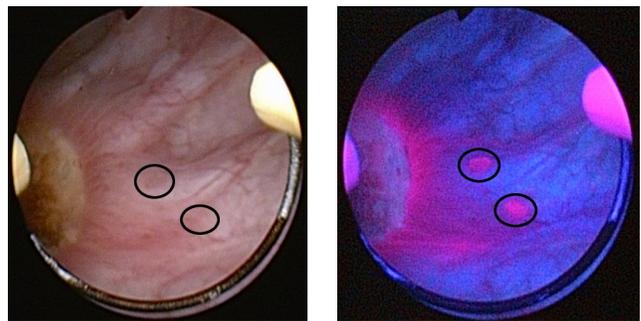


Fig. 1. Example of bladder tumors in normal mode (left) and fluorescence mode (right).

A downside of the approach is the small amount of light left to be detected by the cystoscope's CCD sensor. As a strong electronic amplification is needed for compensation, strong noise is apparent in video sequences of fluorescence cystoscopies.

To overcome this problem, temporal filter techniques can be applied. In this application, real time capable algorithms are mandatory which means that 50 color images must be processed per second at a resolution of 768×576 pixels. An overview of spatial, temporal and spatio-temporal filters is given by Brailean et al. [1]. Most of these filters exhibit complex theory and real time implementations are very difficult to realize. Dubois and Sabri [2] suggest a simple motion compensated temporal recursive filter for which a real time implementation is achievable, but their approach is rather ad hoc and does not guarantee any optimality. Ephraim and Malah introduced a probabilistic minimum mean squared error (MMSE) approach for the spectral amplitude estimation in the domain of speech processing [3] which exhibits very beneficial properties for the remaining noise [4]. Aach adapted their approach to estimate 2D spectral amplitudes in spatial image restoration [5]. Here, we use this concept

We would like to thank Olympus Winter & Ibe GmbH, Hamburg, Germany for funding this project.

to derive a temporal filter directly in the space-time domain rather than in the spectral domain.

The remainder of this paper is organized as follows. In sec. 2, we describe the algorithm of Dubois and Sabri. As we will see in sec. 3, a probabilistic formulation leads to a similar formulation of the filter as Dubois and Sabri suggested and, therefore, justifies their approach. Furthermore, a weighting function is derived which leads to optimal filter results in the MMSE sense. In sec. 4, we present quantitative results of a phantom experiment and qualitative results of filtering a cystoscopic video sequence during a clinical trial. Finally, we conclude the paper with a summary.

2. MOTION COMPENSATED TEMPORAL RECURSIVE FILTER

Let the trivariate function

$$I_n(\vec{x}, t) = I(\vec{x}, t) + n \quad (1)$$

denote the gray value at pixel position $\vec{x} = (x, y)^\top$ of a noise-corrupted image at time t . I represents the ideal noiseless image and n represents zero mean Gaussian distributed noise. For simplicity of notation, we drop the color dimension. If the gray values of an object in the scene do not change with time and both images are noiseless, gray value changes can only be caused by scene motion. This situation is described by the *brightness constancy constraint equation*

$$I(\vec{x}, t) = I(\vec{x} - \vec{d}, t - d_t) \quad (2)$$

where $\vec{d}(x, y, t) = (d_x, d_y)^\top$ is the so called displacement vector field and d_t is the time in which the displacement \vec{d} takes place. Unfortunately, this equation does not hold in the case of fluorescence endoscopy as the images are corrupted by noise. Therefore, the difference of the right hand side and the left hand side of Eqn. (2) is not zero but

$$I_n(\vec{x}, t) - I_n(\vec{x} - \vec{d}, t - d_t) = \tilde{n} \quad (3)$$

governed by a realization \tilde{n} of the noise process contained in both images. For noise reduction Dubois and Sabri suggest a recursive scheme

$$\hat{I}(\vec{x}, t) = \alpha \cdot I_n(\vec{x}, t) + (1 - \alpha) \cdot \hat{I}(\vec{x} - \vec{d}, t - d_t) \quad (4)$$

where \hat{I} is an estimation of the noiseless image signal I and α is a weighting factor. As the image $\hat{I}(\vec{x} - \vec{d}, t - d_t)$ is a motion compensated estimation of the current image $I_n(\vec{x}, t)$, it is called the prediction, and

$$\delta(\vec{x}, t) = I_n(\vec{x}, t) - \hat{I}(\vec{x} - \vec{d}, t - d_t) \quad (5)$$

is defined as the displaced image difference.

Dubois and Sabri further suggest to choose the weighting factor α according to the displaced image difference δ . If it

is low enough (lower than a threshold δ_1) to be explainable by the noise process, a small weighting factor α_1 is chosen such that the prediction has a large influence. If the displaced image difference is high (higher than a threshold δ_2) it is probable that the motion compensation and thus the prediction did not work perfectly. In this case a large weighting factor α_2 is chosen to reduce the influence of the probably imperfect prediction. If the displaced image difference is between δ_1 and δ_2 , α is scaled linearly to give a continuous transition. The piecewise defined function

$$\alpha(\delta) = \begin{cases} \alpha_1, & \text{if } |\delta| \leq \delta_1 \\ \frac{\alpha_1 - \alpha_2}{\delta_1 - \delta_2} |\delta| + \delta_1 \alpha_2 + \delta_2 \alpha_1, & \text{if } \delta_1 < |\delta| \leq \delta_2 \\ \alpha_2, & \text{if } \delta_2 < |\delta| \end{cases} \quad (6)$$

implements these requirements. Note that in this approach *four parameters* ($\alpha_1, \alpha_2, \delta_1, \delta_2$) need to be defined. Fig. 2 gives an example of this weighting function (black solid line).

3. DERIVATION OF OPTIMAL MMSE ESTIMATOR

In this section, we derive the MMSE estimator using a similar reasoning as Ephraim and Malah [3].

The MMSE estimator is defined as the conditional expected value

$$\hat{I}(\vec{x}, t) = E[I_n(\vec{x}, t) | \delta] \quad (7)$$

where \hat{I} , I_n and δ are defined as in sec. 2 and $E[\cdot]$ is the expected value. There are two hypotheses that need to be distinguished: H_0 where the motion compensation worked perfectly, and, therefore, δ is dominated by the noise process. Hypothesis H_1 where the motion compensation failed which means that δ contains noise and, additionally, structural image information. With the theorem of the total probability, Eq. (7) can be rewritten as

$$\hat{I}(\vec{x}, t) = \underbrace{E[I_n | \delta, H_0]}_{\hat{I}(\vec{x} - \vec{d}, t - d_t)} \cdot P(H_0 | \delta) + \underbrace{E[I_n | \delta, H_1]}_{\hat{I}(\vec{x} - \vec{d}, t - d_t) + \delta} \cdot P(H_1 | \delta) \quad (8)$$

where $P(H_0 | \delta)$ and $P(H_1 | \delta)$ are the conditional probabilities of the hypotheses H_0 and H_1 given δ , respectively. Inserting the respective expected values and the displaced image difference δ , one reads

$$\begin{aligned} \hat{I}(\vec{x}, t) &= \hat{I}(\vec{x} - \vec{d}, t - d_t) (P(H_0 | \delta) + P(H_1 | \delta)) + \\ & (I_n(\vec{x}, t) - \hat{I}(\vec{x} - \vec{d}, t - d_t)) \cdot P(H_1 | \delta) \end{aligned} \quad (9)$$

As the probability of a hypothesis plus the probability of its counter hypothesis equals one, the expression can be further simplified

$$\hat{I}(\vec{x}, t) = P(H_1 | \delta) \cdot I_n(\vec{x}, t) + (1 - P(H_1 | \delta)) \cdot \hat{I}(\vec{x} - \vec{d}, t - d_t) \quad (10)$$

which is structurally equal to Dubois and Sabri's suggestion in Eq. (4). One can see that $P(H_1 | \delta)$ is the sought filter curve which corresponds to the weighting function α .

Using the relationships

$$\begin{aligned} P(H_1|\delta) &= 1 - P(H_0|\delta) \\ P(H_0|\delta) &= \frac{p(\delta|H_0) \cdot P(H_0)}{p(\delta)} \quad (\text{Bayes' theorem}) \\ p(\delta) &= p(\delta|H_0) \cdot P(H_0) + p(\delta|H_1) \cdot P(H_1) \end{aligned}$$

the filter curve can be expressed as

$$P(H_1|\delta) = \left(1 + \frac{p(\delta|H_0) \cdot P(H_0)}{p(\delta|H_1) \cdot P(H_1)} \right)^{-1} \quad (11)$$

where $p(\delta)$ is the probability density function (PDF) of δ , and $p(\delta|H_0)$ and $p(\delta|H_1)$ are conditional PDFs. As the regarded noise is based on a quantum process which leads to signal dependent noise, a Poissonian PDF is an appropriate model. But since a Poissonian PDF can be approximated by a Gaussian PDF with a signal dependent σ^2 if sufficient quanta are involved, we assume these PDFs to be zero-mean Gaussians with variances σ_0^2 and σ_1^2 .

Inserting these Gaussians, we obtain

$$P(H_1|\delta) = \left(1 + \frac{P(H_0)\sigma_1}{P(H_1)\sigma_0} \cdot \exp\left(\frac{\delta^2}{2} \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\right) \right)^{-1} \quad (12)$$

In the case of successful motion compensation (H_0 holds), the displaced image difference δ only contains noise. In the case of failed motion compensation (H_1 holds), the displaced image difference does not only contain image noise but also structural image information. Therefore, the corresponding Gaussian's variance σ_1^2 is assumed to be much larger than the noise dominated σ_0^2 and the respective fraction can be neglected in the exponential, yielding

$$P(H_1|\delta) \approx \left(1 + \frac{P(H_0)\sigma_1}{P(H_1)\sigma_0} \cdot \exp\left(\frac{\delta^2}{2\sigma_0^2}\right) \right)^{-1}. \quad (13)$$

The parameter σ_0 can be measured beforehand but, unfortunately, the parameters σ_1 , $P(H_0)$ and $P(H_1)$ remain undetermined. Therefore, we set

$$\lambda = \frac{P(H_0)\sigma_1}{P(H_1)} \quad (14)$$

and rewrite Eq. (13) as

$$P(H_1|\delta) = \left(1 + \frac{\lambda}{\sigma_0} \cdot \exp\left(\frac{\delta^2}{2\sigma_0^2}\right) \right)^{-1}. \quad (15)$$

in which λ is *the only parameter* that needs to be determined empirically. In Fig. 2 three filter curves for $\sigma_0^2 = 1, 49,$ and 225 are depicted.

4. EXPERIMENTS AND RESULTS

The algorithm as described above was implemented in our real time capable clinical demonstrator platform *RealTime-Frame* [6]. A color video stream with a resolution of

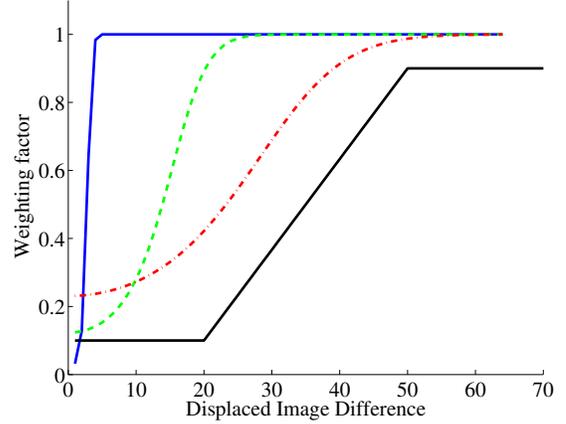


Fig. 2. Different realizations of the filter curve minimizing the MMSE for expected noise variances $\sigma_0^2 = 1$ (blue, solid), 49 (green, dashed), and 225 (red, dash dotted). The parameter λ was fixed to 50. Black: piecewise defined filter curve by Dubois and Sabri.

768×576 pixels at a rate of 50 frames per second can be filtered in real time on our system with two dual-core 2.3 GHz Intel Xeon processors.

For an evaluation of the algorithm's filter performance, a video sequence with a known ground truth signal was necessary. Therefore, a video was acquired with an Olympus Ex-cera II video endoscope in PDD mode. This video shows the inside of a PDD bladder phantom manufactured by Olympus Winter & Ibe GmbH, Hamburg, Germany, without any motion. For this reason, it was possible to average all acquired video frames over time to cancel the noise while preserving the image details:

$$I(\vec{x}) \approx \frac{1}{t_0} \sum_{t=0}^{t_0-1} I_n(\vec{x}, t) \quad (16)$$

Here, $t_0 = 30$ is the number of used images. Subsequently, signal and noise in this sequence could be separated

$$n(\vec{x}, t) \approx I_n(\vec{x}, t) - I(\vec{x}, t). \quad (17)$$

As stated above, noise based on quantum processes is signal dependent and can be approximated by Gaussian models if a large number of quanta is involved. The results of a signal dependent noise analysis are depicted by the blue solid line in Fig. 3

In our first experiment, the proposed algorithm was applied to the motionless video sequence. The filter results on the static data can be regarded as the maximally achievable noise reduction as the static sequence poses an optimal situation to the motion estimator. The resulting noise variances can be seen in Fig. 3 as the green dash-dotted line. As the results, of course, depend on the choice of the parameter λ , it was fixed to a value of 50, which yields visually acceptable results (low detail loss) even in the case of a sequence containing motion.

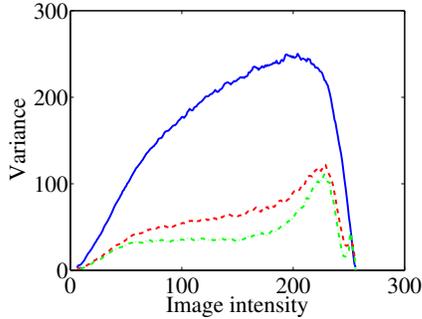


Fig. 3. Noise variance in PDD endoscopy sequence without filtering (blue solid line), with filtering on moved sequence (red dashed line) and with filtering on static sequence (green dash-dotted line).

In a second experiment, the filter performance in presence of a simple motion was evaluated. To this end, again, the dummy sequence was used and an artificial motion was introduced by shifting the image content to neighboring integer pixel positions. We chose integer shifts as shifting to subpixel positions would have affected the noise characteristics because interpolation would then have been necessary. As motion estimator, we used a global linear conformal motion model which exhibits translation, rotation and scaling as degrees of freedom. The motion model was robustly fitted to point correspondences found by block matching. The resulting noise variances can also be seen in Fig. 3 (red dashed line). As expected, the noise reduction was less effective compared to the static sequence, but still a substantial noise reduction could be achieved.

The peaks on the right hand side of the remaining noise variances of the filtered sequences in Fig. 3 can be explained by the effects of imperfect motion estimation. In these cases, the edges of dark blood vessels were not perfectly matched and therefore the difference δ increased and the filter strength was reduced.

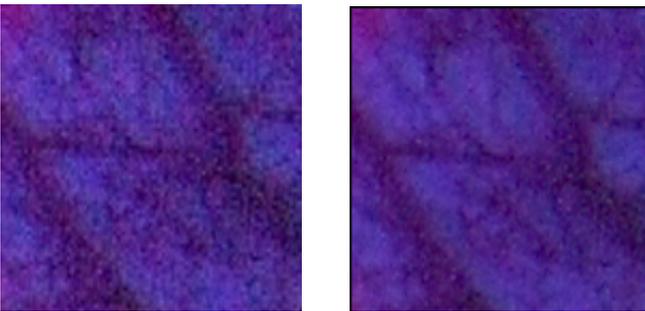


Fig. 4. Image from endoscopy sequence without any filtering (left) and same image after application of the proposed filter (right). Image details are widely preserved or even enhanced (e. g. blood vessels).

In a first clinical trial, our system was placed next to the original endoscopy monitor the physician could directly compare both images. The physician commented positively on the ef-

fects of the filter algorithm. Qualitative results are depicted in Fig. 4. On the left hand side, an original unfiltered frame from a human bladder examination and, on the right hand side, the same frame after application of the proposed filter can be seen. Despite scene motion, noise is substantially reduced while image details are widely preserved or even enhanced (blood vessels).

5. CONCLUSION

In this paper, we have given a short introduction to PDD cystoscopy and explained the presence of strong noise in this type of video stream. As countermeasure, we have derived a filter framework for effective noise reduction adapting a spectral estimation approach from speech processing. Thereby, we have developed a probabilistic foundation for Dubois and Sabri's temporal filter. Furthermore, we have derived a weighting function which gives optimal results in the MMSE sense. If the signal dependent system noise is known, the proposed function needs only one parameter to be defined. For the evaluation of the algorithm, we have implemented it on our real time capable software framework *RealTime-Frame*. We have measured and compared the noise variances in a unfiltered dummy sequence and in filtered sequences without and with artificial motion. During a clinical trial, the physician commented positively on the improved image quality.

6. REFERENCES

- [1] J.C. Brailean, R.P. Kleihorst, S. Efstratiadis, A.K. Katsaggelos, and R.L. Lagendijk, "Noise reduction filters for dynamic image sequences: a review," *Proc. IEEE*, vol. 83, no. 9, pp. 1272–1292, 1995.
- [2] E. Dubois and S. Sabri, "Noise reduction in image sequences using motion-compensated temporal filtering," *IEEE Tr. Comm.*, vol. 32, no. 7, pp. 826–831, 1984.
- [3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE ASSP*, vol. 32, no. 6, pp. 1109–1121, 1984.
- [4] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE SAP*, vol. 2, no. 2, pp. 345–349, 1994.
- [5] T. Aach, "Spectral transform-based nonlinear restoration of medical images: algorithms and comparative evaluation," in *SPIE*, 1999, vol. 3646, pp. 270–280.
- [6] S. Gross and T. Stehle, "RealTimeFrame – A Real Time Processing Framework for Medical Video Sequences," *Acta Polytechnica, Journal of Advanced Engineering*, vol. 48, no. 3, pp. 15–19, 2008.