A Frequentistic and a Bayesian Approach for Optimal Optical Filter Design.

Miro Taphanel

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
miro.taphanel@kit.edu

Technical Report IES-2013-11

Abstract: This report discusses three merit functions to optimize optical interference filter coatings. The applications of these filters are intentionally optical 3D sensors, e.g. a chromatic confocal triangulation sensor. Optimizing these optical filters is done by minimizing the measurement uncertainty of the sensor. The measurement task is handled as a parameter estimation problem and the sensor is considered as a physical experiment. As part of the experimental design, the optical filters are optimized to achieve measurements with lower uncertainty. The first merit function is based on a frequentistic statistic utilizing the *Cramér-Rao* lower bound. An example is used to point out disadvantages and two alternative merit functions are proposed. Instead of a lower bound, the other merit functions incorporate a specific estimator function.

1 Introduction

Designing a sensor from scratch offers many degrees of freedom. The process is equal to setup an experiment and fixing all the design variables in the sense of an optimal experimental design. In literature [HK05],[Bos07], [Ber85],[CV95] experimental design is a well-studied topic. The basic idea is to apply estimation theory to model the outcome of an experiment. On top of this model optimality criteria are defined, which quantify the performance of the experiment. Finally, using these criteria as merit functions in an optimization framework will lead to improved experimental designs. In [Bos07],[VAdDVDvdB02] it is proposed to utilize the *Cramér-Rao* lower bound to quantify the variance of the experimental outcome. The *Cramér-Rao* lower bound is a general lower bound of the variance of an arbitrary estimator function [Bos07]. Because the purpose of this research is to optimize a measurement sensor, the variance of an estimator function is of special interest. According to [fS04], the uncertainty of a measurement is quantified by variances or standard deviations and the measurement itself is only an estimate of the value of the measurand. The main advantage of the *Cramér-Rao* lower bound is its compact closed form expression. Unfortunately, the *Cramér-Rao* lower bound implicit linearizes the physical model for a given set of design variables. This report emphasizes the resulting drawbacks for oscillating non-linear models. An example similar to [VDBCT03] is presented. To overcome this problem it is proposed to use a specific estimation function instead of a lower bound. For this purpose [MVDBB94] proposed to use the variance of a least square estimator. However, the sensor model had to be linearized. A general approach is Bayesian experimental design [HM13], [CV95],[VDBCT03], [Ber85]. In [HM13] the application of Bayesian experimental design is shown for nonlinear models. They optimize an experiment based on a merit function utilizing the *Kullback-Leibler* divergence. The *Kullback-Leibler* divergence is used as distance measure between the posterior and the priori and quantifies the information gain made by an experiment. The idea was originally proposed by [Lin56] and is derived from *Shannon* information theory. The principle approach was generalized [Lin72],[CV95] to allow other utility functions than the entropy as information measure. In section 3.3 this approach is used in combination with a variance like utility function. However this approach lead to experimental design, which are optimal on average. As an alternative in section 3.4 a merit function for experimental design is proposed, which optimizes always the worst case. In this case there is no risk that some working points of the experiment have higher uncertainty for the benefit for others.

2 Sensor Model

This section provides a rough sensor model. For simplification details are neglected but can be found in [THB13]. The intention of this section is to clarify the notation and the application. In the next section estimation theory is applied to the provided model.

The interference filters are optimized for a chromatic confocal triangulation (CCT) [TB12] sensor. In principle, interference filters can realize arbitrary transmission characteristics by customized thin film layer stacks. The scope of this research is to optimize the sensor by adjusting the thicknesses of these thin film layers, which in turn change the filter transmissions. Assume a CCT sensor with six filters corresponding to six camera channels. The gray values of each channel are organized in a vector and denoted as $\mathbf{g} = (g_1, \ldots, g_6)^\top$. Each filter is determined by its thin film layer stack. The characteristic thicknesses of each layer are organized as a parameter vector $\mathbf{p}_i$ and the index i specifies the corresponding optical filter. For simplification all filters are summarize in one long vector $\mathbf{p}$. In experimental design these parameters are sometimes called design variables.

The measurement procedure of a CCT sensor is to estimate a height, which is optical encoded by a wavelength λ, based on the gray values $\mathbf{g}$. Apart of a nonlinear relationship, height and wavelength λ are equivalent and instead of the height, λ is used as the parameter of interest. Assuming an arbitrary estimation function $f(.)$ the normal working procedure of a CCT senor can be formalized as:

$$\hat{\lambda} = f(\mathbf{g}; \mathbf{p}),$$

with the target to estimate the corresponding wavelength. In estimation literature the parameters $\mathbf{p}$ are denoted as nuisance parameter, because they are not of interest. In experimental design these parameter are the adjustment screws to gain better performance.

A requirement to apply powerful estimation functions, like the *Maximum Likelihood* estimation, is to specify the distribution of the measurements. The dominant non-systematic error source in the CCT sensor is the photon noise of the involved camera. For large number of photons the *Poisson* distribution can be approximated by the normal distribution. For this case the six channel camera gray value $\mathbf{g}$ is modeled as random variable G:

$$\begin{aligned}
&\mathrm{E}\{G\} = \mathbf{g}_\mu(\lambda; \mathbf{p}), \mathbf{g}_\mu : \mathbb{R} \to \mathbb{R}^6, \lambda \mapsto \mathbf{g} \\
&G \sim \mathcal{N}\left(\mathbf{g}_\mu(\lambda; \mathbf{p}), \mathrm{diag}(\sigma_1^2(\lambda; \mathbf{p}), \ldots, \sigma_6^2(\lambda; \mathbf{p}))\right), \text{with} \\
&\sigma(\lambda, \mathbf{p}) = \sigma_d + k\mathbf{g}_\mu(\lambda; \mathbf{p}), \sigma : \mathbb{R} \to \mathbb{R}^6, \lambda \mapsto \sigma = (\sigma_1, \ldots, \sigma_6)^\top \\
&p(\mathbf{g}|\lambda, \mathbf{p}) = \prod_{i=1}^{6} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\left(\frac{g_i - g_{\mu,i}(\lambda;\mathbf{p})}{\sigma_i}\right)^2}
\end{aligned} \tag{2.1}$$

The sensor model $\mathbf{g}_\mu(\lambda; \mathbf{p})$ defines the expectation value of G. In [Bos07] this sensor model is called expectation model. The random variable G is assumed to be normal like distributed and each of the six camera channels is assumed to be statistically independent. The independence property results in a diagonal covariance matrix. The variance of each camera channel is a function of the sensor model again, to realize an approximation of the Poisson distribution.

For the Bayesian framework the deterministic parameter λ is considered as a random variable Λ. A non-informative a priori probability density function is assumed:

$$p(\lambda) = \begin{cases} \frac{1}{\lambda_{\max} - \lambda_{\min}}, & \text{if } \lambda_{\text{min}} \leq \lambda \leq \lambda_{\text{max}} \\ 0, & \text{else,} \end{cases}$$

which just expresses the knowledge that the wavelength will be within certain boundaries. Using the Bayes' theorem, the a posteriori probability density function is given by:

$$\begin{aligned} p(\mathbf{g}|\mathbf{p}) &= \int p(\mathbf{g}|\lambda, \mathbf{p}) p(\lambda) \mathrm{d}\lambda \\ p(\lambda|\mathbf{g}, \mathbf{p}) &= \frac{p(\mathbf{g}|\lambda, \mathbf{p}) p(\lambda)}{p(\mathbf{g}|\mathbf{p})} \\ &= \begin{cases} \frac{p(\mathbf{g}|\lambda, \mathbf{p})}{\int_{\lambda_{\min}}^{\lambda_{\max}} p(\mathbf{g}|\lambda, \mathbf{p}) \mathrm{d}\lambda}, & \text{if } \lambda_{\text{min}} \leq \lambda \leq \lambda_{\text{max}} \\ 0, & \text{else.} \end{cases} \end{aligned}$$

3 Optimizing the Experimental Sensor Design

In this section merit functions are derived to optimize the sensor performance. Optimizing the performance of such a sensor aims to minimize the measurement uncertainty. According to [fS04] the measurement uncertainty is defined as standard deviation (or variance) of the measurement result, while the measurement is only an estimation of the true value. Because the measurement process is an estimation procedure, the optimization tries to minimize the variance of the estimation. The following subsections define different design criteria, which propose optimal design parameters $\mathbf{p}^\star$ for optimal experimental design settings.

3.1 Cramér-Rao Lower Bound Approach

The *Cramér-Rao* lower bound is a fundamental lower bound for the variance of any estimator. Because the lower bound is a function of the experimental design parameters, too, it is a easily accessible way to improve an experiment. The assumption behind this approach is that estimators are available, which reach this lower bound at least asymptotically. A famous example is the *Maximum Likelihood* estimator[Bos07] (p. 81). According to [Bos07] the *Cramér-Rao* lower bound for normal distributed observations is defined as:

$$\mathrm{Var}\{f(\mathbf{g}; \mathbf{p})\} \geq \left(\frac{\partial \mathbf{g}^\top(\lambda, \mathbf{p})}{\partial \lambda} \mathbf{C}^{-1} \frac{\partial \mathbf{g}(\lambda, \mathbf{p})}{\partial \lambda} \right)^{-1},$$

with a covariance matrix C. In the CCT sensor application only one parameter λ is of interest, this scalar variance measure can directly be used as a merit function to optimize the experimental design. Because a sensor is only as good as its worst working point, a *Minimax* optimization is proposed:

$$\mathbf{p}^{\star} = \arg\min_{\mathbf{p}} \max_{\lambda} \mathrm{Var}\left\{ f\left(\mathbf{g}(\lambda, \mathbf{p})\right)\right\},$$

which concentrates on minimizing the highest variance whithin the measurement range.

As comparison in [THB13] several merit functions were presented. To link this result to these, a slightly different noise model (2.1) is assumed. For this comparison the covariance matrix $\mathrm{Cov} = \sigma^2\mathbf{I}$ is modeled with constant standard deviation σ and identity matrix $\mathbf{I}$. In this case the *Cramér-Rao* lower bound can be expressed as:

$$\begin{aligned}\mathrm{Var}\{f(\lambda)\} &= \left(\frac{\partial \mathbf{g}^{\top}(\lambda, \mathbf{p})}{\partial \lambda}(\sigma^2\mathbf{I})^{-1}\frac{\partial \mathbf{g}(\lambda, \mathbf{p})}{\partial \lambda}\right)^{-1} \\ &= \frac{1}{\sigma^2}\left(\left\|\frac{\partial \mathbf{g}(\lambda, \mathbf{p})}{\partial \lambda}\right\|_2^{-1}\right)^2, \end{aligned} \tag{3.1}$$

which is identical to the proposed ”sensitivity” merit function in [THB13] and reflects the result in a different light.

Unfortunately, this kind of experimental optimization will fail due to the non-linearity of the CCT sensor model. The model $\mathbf{g}(\lambda, \mathbf{p})$ is highly non-linear and has in particular an oscillating character. Optimizing only the lower bound of the estimation variance will lead to an ill-posed estimation problem. The well posed property will be lost, because the oscillating character of $\mathbf{g}(\lambda, \mathbf{p})$ will cause ambiguities. To clarify the problem, an example is provided in the next section.

3.2 Example - Effect of Non-Linear Models

The following example is intent to emphasize the problem of a non-linear CCT sensor model. Especially, the oscillating function character leads to ambiguities and causes the estimation problem to be ill-posed. The example is adapted from [VDBCT03]. Instead of investigating the CCT sensor model an simplified sensor model is assumed:

$$g_{\mu}(\lambda, \mathbf{p}) = \frac{1}{2}\sin\left(p_1(\lambda - p_2)\right) + \frac{1}{2}, \tag{3.2}$$

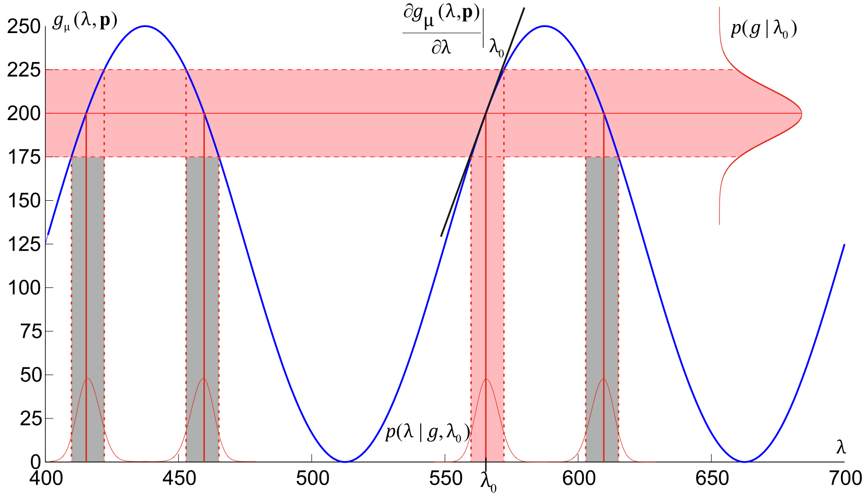

Figure 3.1: Blue graph depicts the non-linear (sinus like) relationship between the gray value $g_\mu(\lambda, \mathbf{p})$ and λ. Furthermore, the normal distribution $p(\mathbf{g}|\lambda_0)$ is depicted and the gradient at λ_0 as part of the *Cramér-Rao* lower bound. Finally, the posteriori probability density function is illustrated, too. The posteriori shows four peaks with equal probability. However, this ambiguity is not recognized by the gradient used to calculate the *Cramér-Rao* lower bound. The example is adapted from [VDBCT03].

with clear oscillating character. The model is depicted in Fig. 3.1 as blue graph. This model lead to an estimation process which is ill-posed due to ambiguities. Assuming a measurement (observation) of $\mathbf{g} = 200$, there is no evidence to prefer one of the four estimates: $\hat{\lambda} \in \{415, 460, 565, 610\}$. The task of a experimental design is to remove the ill-posed property and too ensure measurements with low uncertainty. The key idea is, that this can be done in parallel if the current estimation variance is minimized. Ambiguities in the estimation process increase the uncertainty of the estimate and thus the variance of the estimator.

Assume that the frequency p_1 and the offset p_2 in the example model (3.2) would be adjustable design parameters of the experiment. Then, an optimal solution for a setup with well posed estimation process is shown in Fig. 3.2. The depicted solution is optimal, because every higher frequency p_1 would introduce an ambiguity. On the other side, a lower frequency would decrease the gradient $\partial g_\mu / \partial \lambda$ and according to the *Cramér-Rao* lower bound increase the variance of the estimation. The step between the result depicted in Fig. 3.1 and the proposed preferred

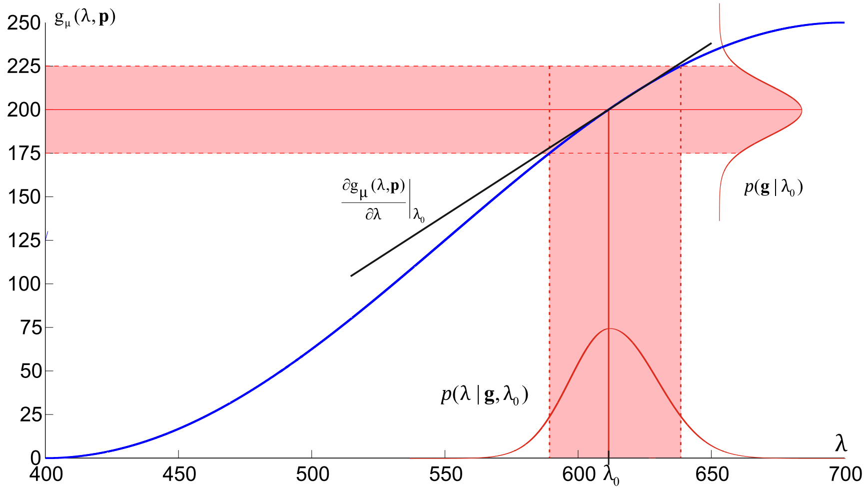

Figure 3.2: Blue graph shows a sensor model $g(\lambda)$ in optimal experimental settings. The estimation problem is well-posed and the uncertainty for this case is minimal.

result in Fig. 3.2 is optimizing the experimental design. However, the *Cramér-Rao* lower bound approach will lead to a contrary result. According to equation (3.1), the lower bound involves a sensor model gradient. With view to the example model (3.2) an optimized design would increase the frequency p_1 to infinity, because:

$$\frac{\partial \mathbf{g}(\lambda, \mathbf{p})}{\partial \lambda} = \frac{1}{2}\cos(p_1(\lambda - p_2))p_1 \leq \frac{1}{2}p_1.$$

This shows clearly that ambiguities are not recognized by the local gradient.

In the following two sections experimental design approaches are presented, which incorporate a specific estimation function. If e.g. an *Maximum-a-Posteriori Probability* (MAP) estimator is used, the estimate is just the maximum of the posteriori probability density $p(\lambda|g, \lambda_0)$. As depicted in Fig. 3.1 the posteriori probability density function consists of four asymmetric gaussian like distributions. These four peaks contain the information of an increased measurement uncertainty, caused by ambiguities due to the non-linear sensor model. Utilizing the MAP estimator variance will prevent the experimental design optimization to turn into an ill-posed problem.

3.3 Bayesian Experimental Design

The last two sections show that the *Camér-Rao* lower bound variance measure is not suitable for non-linear models. This section overcomes the disadvantages of a lower bound by involving a estimation function. Restricting to a specific estimation function allows to access the variance without approximation. As shown in example 3.1, ambiguities heavily increase the estimation variance. Minimizing this variance in an experimental design optimization will prevent ambiguities in the estimation process.

The idea to incorporate a concrete estimation function in experimental design does not justify to change over from a frequentistic to a Bayesian approach. It's rather a free decision of the author. In [CV95] a general approach of Bayesian experimental design was presented. An experimental design is defined by the design variables $\mathbf{p}$ and observations $\mathbf{g}$ which will be made in the experiment. Based on $\mathbf{g}$ an estimation function $\hat{\lambda} = f(\mathbf{g}; \mathbf{p})$ estimates the unknown parameter of interest λ. Then, the best Bayesian experimental design is given by [CV95]:

$$\mathbf{p}^\star = \arg\min_{\mathbf{p}} \min_{\mathbf{f}\in\mathcal{F}} \int\int u(f, \lambda, \mathbf{p}, \mathbf{g}) p(\lambda|\mathbf{g}, \mathbf{p}) p(\mathbf{g}|\mathbf{p}) \mathrm{d}\lambda \mathrm{d}\mathbf{g}. \tag{3.3}$$

The utility function $u(f, \lambda, \mathbf{p}, \mathbf{g})$ reflects the purpose of the experiment and with the idea of a variance measure it is chosen to $u(f, \lambda, \mathbf{p}, \mathbf{g}) = (\lambda - \hat{\lambda})^2$. The double minimization takes into account, that both, a suitable estimation function f out of a set of estimation function $\mathcal{F}$ and the best design parameters $\mathbf{p}$ must be chosen. Suitable estimation function are e.g. the *Bayesian* estimator:

$$\hat{\lambda} = \int \lambda p(\lambda|\mathbf{g}, \mathbf{p}) \mathrm{d}\lambda$$

and the *Maximum a posteriori* (MAP) estimator:

$$\hat{\lambda}_{\mathrm{MAP}} = \arg\max_{\lambda} p(\lambda|\mathbf{g}, \mathbf{p}).$$

Without prove, the MAP estimator is preferred, because the non-linear model will cause an asymmetric posterior probability density function which will cause a bias for the Bayesian estimator. Although, the Bayesian estimator is proven to have the lowest variance [Ber85](p.136).

3.4 Worst Case Experimental Design

The approach of the Bayesian experimental design (section 3.3) contains a hidden risk. The integral over λ causes an averaging over all possible working points.

Thus an experimental design can be improved by increasing the measurement uncertainty of a single working point for the benefit for others. However, a sensor is only as good as its worst working point and this behavior is undesirable. In literature, the worst case optimization in combination with experimental design is rarely studied. A related idea *Maximum Mean Squared Error* optimization was proposed by [SSW89] and [SWMW89]. In [Coh96],[SHL12] an similar idea was discussed. As a side note, the following formulation of a merit function is neither purely frequentistic nor Bayesian. For a given working point λ_0, the squared difference of an estimator function $\hat{\lambda} = f(\mathbf{g})$ is given by:

$$u(f, \lambda_0, \mathbf{p}, \mathbf{g}) = (\hat{\lambda} - \lambda_0)^2,$$

as a function of the observations $\mathbf{g}$ and its corresponding random variable G. According to [Coh96] the expected mean squared error (MSE) is given by:

$$\mathrm{E}_{\mathrm{MSE}}\{G\} = \int (\hat{\lambda} - \lambda_0)^2 p(\mathbf{g}|\lambda_0, \mathbf{p}) \mathrm{d}\mathbf{g}.$$

This equation evaluates the expected MSE at the working point λ_0. The best worst case experimental design is then given by:

$$\mathbf{p}^\star = \arg\min_{\mathbf{p}} \max_{\lambda_0} \int (\hat{\lambda} - \lambda_0)^2 p(\mathbf{g}|\lambda_0, \mathbf{p}) \mathrm{d}\mathbf{g}.$$

In contrast to (3.3) the optimization of the selected estimation function was neglected.

4 Conclusion

The research points out, that the *Cramér-Rao* lower bound implicit linearizes a sensor model. Using the lower bound to optimize the experimental setup for nonlinear models is problematic. For the application to optimize interference filters for a CCT sensor, the optimized experimental design results in a ill-posed estimation task. To avoid this problem an alternative approach is proposed. Specifying a concrete estimator, the estimation variance can directly be minimized. Open questions are an experimental validation with a comparison between the Bayesian experimental design and the proposed worst case experimental design. In an former publication [THB13] the problem of an ill-posed estimation process was avoided by an additional merit function. A comparison with this approach would be interesting, too. Another open question is the selection of a suitable estimation function.

A topic that was not tackled is the numerical realization in an optimization framework. Due to the non-linearities, the overall optimization problem is highly non-convex. The found optimized experimental design will be a local optimum with high probability. For this reason, the calculation complexity will influence the quality of experimental design, too. For the application itself, the calculation speed is of great importance.

Bibliography

[Ber85] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, August 1985.

[Bos07] Adriaan van den Bos. *Parameter Estimation for Scientists and Engineers*. John Wiley & Sons, August 2007.

[Coh96] David A. Cohn. Neural network exploration using optimal experiment design. *Neural Networks*, 9(6):1071–1083, August 1996.

[CV95] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, August 1995. ArticleType: research-article / Full publication date: Aug., 1995 / Copyright © 1995 Institute of Mathematical Statistics.

[fS04] International Organization for Standardization. *Guide to the expression of uncertainty in measurement (GUM)-Supplement 1: Numerical methods for the propagation of distributions*, volume ISO draft guide DGUIDE99998. International Organization for Standardization, Geneva, 2004.

[HK05] Klaus Hinkelmann and Oscar Kempthorne. *Design and Analysis of Experiments: Advanced experimental design*. Wiley, April 2005.

[HM13] Xun Huan and Youssef M. Marzouk. Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317, January 2013.

[Lin56] D. V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, December 1956. ArticleType: research-article / Full publication date: Dec., 1956 / Copyright © 1956 Institute of Mathematical Statistics.

[Lin72] D. V. Lindley. *Bayesian Statistics, A Review*. SIAM, January 1972.

[MVDBB94] M. A O Miedema, A. Van Den Bos, and A. H. Buist. Experimental design of exit wave reconstruction from a transmission electron microscope defocus series. *IEEE Transactions on Instrumentation and Measurement*, 43(2):181–186, 1994.

[SHL12] Antti Solonen, Heikki Haario, and Marko Laine. Simulation-based optimal design using a response variance criterion. *Journal of Computational and Graphical Statistics*, 21(1):234–252, 2012.

[SSW89] Jerome Sacks, Susannah B. Schiller, and William J. Welch. Designs for computer experiments. *Technometrics*, 31(1):41, February 1989.

[SWMW89] Jerome Sacks, William J. Welch, Toby J. Mitchell, and Henry P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423, November 1989. Mathematical Reviews number (MathSciNet): MR1041765; Zentralblatt MATH identifier: 0955.62619.

[TB12] M. Taphanel and J. Beyerer. Fast 3D in-line sensor for specular and diffuse surfaces combining the chromatic confocal and triangulation principle. In *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International*, pages 1072 –1077, May 2012.

[THB13] Miro Taphanel, Bastiaan Hovestreydt, and Jürgen Beyerer. Speed-up chromatic sensors by optimized optical filters. pages 87880S–87880S, May 2013.

[VAdDVDvdB02] S. Van Aert, A.J. den Dekker, D. Van Dyck, and A. van den Bos. Optimal experimental design of STEM measurement of atom column positions. *Ultramicroscopy*, 90(4):273–289, April 2002.

[VDBCT03] Jojanneke Van Den Berg, Andrew Curtis, and Jeannot Trampert. Optimal nonlinear bayesian experimental design: an application to amplitude versus offset experiments. *Geophysical Journal International*, 155(2):411–421, 2003.