RNN-accelerated Experimental Design for Chromatic Confocal Measurement

Ding Luo

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

Technical Report IES-2016-02

Abstract: With decades of research and development, confocal microscopes have been the work horse of scientific and industrial 3D measurement. However, due to its requirement for axial scanning, its range of application is limited by its slow measurement speed. Chromatic confocal measurement systems have been developed to eliminate the need for mechanical scanning. Nevertheless, they are still bottle-necked by the transfer and processing of densely sampled spectral data. In this article, Bayesian experimental design is applied to the chromatic confocal measurement scheme, allowing for more efficient spectral sampling. Recurrent neural network (RNN) is trained to approximate full Bayesian experimental design with much less computation. Simulations have demonstrated that experimental design approximated by RNN provides better results than an equidistant sampling scheme and performance close to full Bayesian experimental design.

1 Introduction

Confocal microscope has been widely applied in various fields due to its superior resolution and unique depth-discerning capability [HWS81]. In a conventional confocal system, a monochromatic point source is projected into the object space for illumination and the returning light is reflected to a pinhole which guarantees that light gets fully transmitted to the detector only when the object is in focus. In order to retrieve the depth information, the system has to be shifted axially with respect to the object while the detector records intensity measurements at different locations. This scanning generates a Gaussian-like

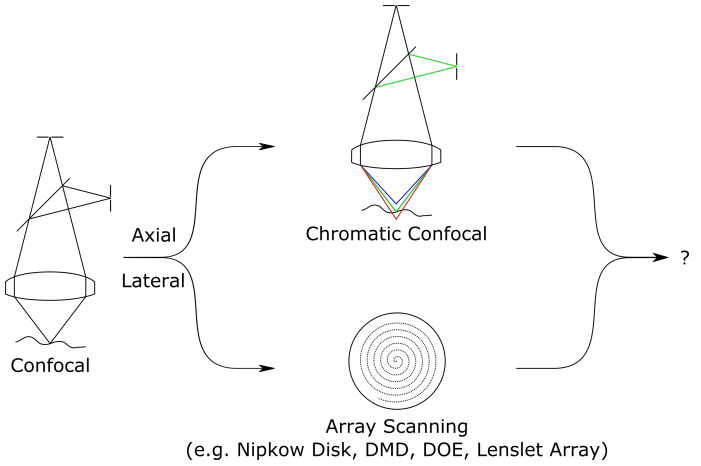

Figure 1.1: Confocal microscope requires both axial and lateral scanning.

signal which can be utilized to retrieve the depth of the object. In order to get a full 3D measurement of the object, the system has to be scanned also laterally. Confocal measurement is slow due to its requirement for axial and lateral scanning. Various technologies have been developed to accelerate these processes, such as chromatic confocal microscope [MPPQ84] and various array scanning microscopes [EAH02, DLY+13]. However, two fundamental problems remain unsolved. In the axial direction, when we are dealing with confocal or chromatic confocal signal, we are always trying to locate the location of a quasi-Gaussian peak which directly reflects the location/depth of the object under measurement. According to estimation theory the width of the peak is directly related to the sensitivity of the system, and in this case people are always trying to build a system with a peak as narrow as possible in order to increase the sensitivity. Nevertheless, a narrower peak means that the information about the object is only contained within a tiny subset of support where all the rest of the signal are simply zeros. To measure the complete signal with such sparsity like what's typically done in a chromatic confocal system with spectrometer is highly inefficient. The second fundamental problem is related with the density of measurement. Light which is not focused on the object will spread across adjacent area of a single measurement point, limiting the minimum distance between two simultaneous lateral measurement points that prevents crosstalk. This article aims to tackle the first problem through simulations of Bayesian experimental design and its accelerated approximation based on Recurrent Neural Network.

2 Chromatic Confocal Measurement Model

The target of chromatic confocal measurement is to retrieve the depth of the measurement position via the location of the Gaussian-like peak. From the point of view of parameter estimation, the canonical way to do this is to build a measurement model and apply Bayesian inference on the parameters of interest. The measurement model is composed of two parts, the signal model and the noise model. The signal model describes the relationship between an ideal signal, or expectation of the signal, and the parameters to be estimated. And the noise model represents the amount of noise added to the ideal model.

In the case of chromatic confocal measurement, we assume that the signal is a Gaussian function expressed by the following equation:

$$\hat{y} = \theta_1 e^{-\frac{(\lambda-\theta_2)^2}{2\sigma^2}}$$

where θ_1 represents the amplitude of the signal and θ_2 represents the location of the signal. θ_1 is mainly determined by the reflectance of the object and θ_2 reflects the axial position of the object. λ denotes the wavelength. σ represents the width of the Gaussian-shaped chromatic confocal signal and is determined by properties of the optical system such as the numerical aperture. Assuming normally distributed noise, the complete model is expressed as a normal distribution over the combination of signal and noise:

$$y \sim \mathcal{N}(\hat{y}, \sigma_n) \tag{2.1}$$

where σ_n describes the variance of the noise and is mainly determined by the camera.

Based on Bayes theorem, the parameter estimation task is relatively straightforward by calculating the posterior probability distribution of the parameters based on the measurement model. In this case, the parameters of interest are $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$, where θ_1 contains information of the object texture and θ_2 contains the depth information. The posterior is proportional to the product of prior and likelihood. Without any prior knowledge, the prior distribution is considered to be flat across the valid support so that all parameter values are equally possible when no measurements are made. The likelihood comes directly from the measurement model, as shown in Eq. (2.1). Therefore, the posterior distribution can be calculated up to a certain scale factor:

$$\begin{aligned} p(\boldsymbol{\theta}|y) &= \frac{p(\boldsymbol{\theta})p(y|\boldsymbol{\theta})}{p(y)} \\ &\propto p(\boldsymbol{\theta})p(y|\boldsymbol{\theta}) \end{aligned}$$

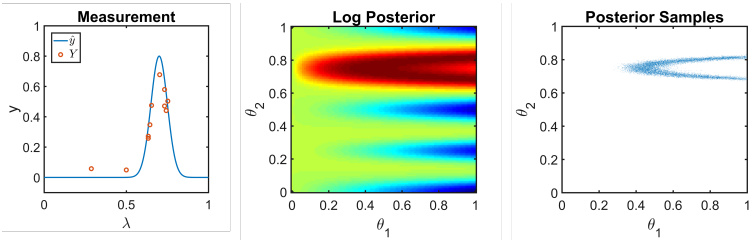

Figure 2.1: Posterior sampling after measurements are made.

In practice, calculating the posterior distribution with certain resolution is often computationally prohibitive, and therefore people have been using sampling techniques such as Markov-Chain Monte Carlo (MCMC) method. In our case, an ensemble sampler which is affine-invariant [GW] is utilized for drawing the posterior samples. Once samples are drawn from the posterior distribution, the estimation becomes trivial by calculating the average of all samples.

Figure 2.1 demonstrates the procedure of posterior sampling for chromatic measurement through simulation. In the left figure, the signal to be measured is denoted by $\hat{y}$ and simulated measurements with normally distributed noise are contained in Y . The middle figure illustrates the posterior probability distribution of the parameters to be estimated and the right figure shows samples drawn from such distribution.

The Bayesian framework has two major advantages for parameter estimation. Firstly, the uncertainty of the estimation can be easily derived by calculating the variance of the samples. Secondly, the posterior distribution of the parameter allows for the selection of optimal measurement location in the next measurement through Bayesian experimental design, as will be discussed in the next section.

3 Bayesian Experimental Design

Bayesian experimental design is the subject of making decisions under uncertainties with limited resource. In our case of measuring a chromatic confocal signal, conventional systems utilize a spectrometer which disperse various wavelength onto hundreds of pixels. A major drawback for such approach is that the transfer of the intensity data can be quite slow. Additionally, in the case of multi-point chromatic confocal system, the application of multiple spectrometers is often prohibitive, due to either cost or mechanical constrains. Therefore, wavelength scanning of the light source is used instead to acquire the chromatic signal. Nev-

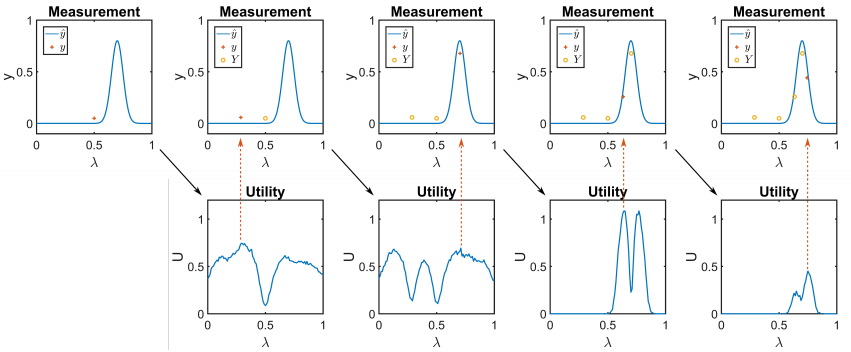

Figure 3.1: Adaptive measurement of a chromatic confocal signal. First row: each measurement step. Second row: utility function after each measurement step.

ertheless, such process can be time-intensive depending on the scanning speed of the light source.

Instead of an equidistant measuring scheme, Bayesian experimental design allows for an adaptive measuring scheme, where location for new measurement is determined by measurements already conducted. For example, With a tunable light source whose wavelength can be tuned programmably, suppose that we have measured intensities of five wavelengths, the question that Bayesian experimental design tries to answer is which wavelength we should look at next so that estimation could be made most efficiently in the end.

The profit generated by measurement at a certain wavelength is described by a utility function over the design space. There are various different utility functions which focuses on different aspects of the design. Here we will take the Kullback-Leibler divergence between the posterior and the new posterior with the new measurement. The utility function is expressed as the expectation of this KL divergence under the posterior predictive distribution [Rya03]:

$$
\begin{aligned}
U(d) &= \mathbb{E}_{y|d}[D_{\mathrm{KL}}(p(\boldsymbol{\theta}|Y,y,d)\|p(\boldsymbol{\theta}|Y))] \\
&= \iint p(\boldsymbol{\theta}|Y,y,d)\log\frac{p(\boldsymbol{\theta}|Y,y,d)}{p(\boldsymbol{\theta}|Y)}\mathrm{d}\boldsymbol{\theta}p(y|Y,d)\mathrm{d}y \\
&= \iint p(\boldsymbol{\theta}|Y)p(y|\boldsymbol{\theta},d)\big[\log p(y|\boldsymbol{\theta},Y,d) \\
&\quad - \log[\int p(\boldsymbol{\theta}|Y,d)p(y|\boldsymbol{\theta},Y,d)\mathrm{d}\boldsymbol{\theta}]\big]\mathrm{d}\boldsymbol{\theta}\mathrm{d}y
\end{aligned}
$$

$$\approx \frac{1}{N}\sum_{i=1}^{N}\Big\{\log p(y^i|\boldsymbol{\theta}^i,d) - \log\big[\frac{1}{M}\sum_{j=1}^{M} p(y^i|\boldsymbol{\theta}^{ij},d)\big]\Big\}$$
$$\{\boldsymbol{\theta}^i\}\cup\{\boldsymbol{\theta}^{ij}\}\sim p(\boldsymbol{\theta}|Y), \{y^i\}\sim p(y|\boldsymbol{\theta}^i,d)$$

where d represents possible design, i.e. next wavelength to be measured. Calculation of the double integral for this utility function cannot be conducted analytically and therefore is solved by a nested Monte Carlo approximation using posterior samples drawn for parameter estimation.

Finally, the task is to find the d^* which maximizes the utility function above. Although there are stochastic optimization techniques for such problems, people typically calculate the utility functions for a grid of design points and take the maximum one. Notice that this approach is based on the so-called myopic design. It means that only one further step is considered based on current situation. This does not guarantee true optimal design for an experiment with multiple measurements, but in general works very well as a greedy method.

As an example, Figure 3.1 demonstrates the adaptive measurement of a chromatic confocal signal. The first row shows the signal to be measured and the corresponding measurement in each step. In these graphs, $\hat{y}$ denotes the signal to be measured, y represents new measurement in each step and Y contains all measurements conducted. The second row shows the utility function over the design space in each step. In this example, measurement starts by recording intensity of the wavelength in the middle. Based on the measurement result, parameter estimation is conducted and the utility function over all wavelengths is calculated. In the next step, intensity is measured at the wavelength which has the largest utility value.These two steps can be repeated multiple times until utility functions for all positions drops to zero, indicating that new measurements do not bring any additional information. The wavelength is normalized to a range from zero to one as the calculations are all based on simulations.

Figure 3.2 shows the comparison between equidistant measurement scheme and adaptive measurement scheme based on Bayesian experimental design. As seen from the posterior samples, with same number of measurement steps, the adaptive approach typically generates much concentrated samples, indicating less uncertainty for parameter estimation. The reason is that the adaptive approach tends to make new measurements at locations where more information is expected to be gained about the parameters.

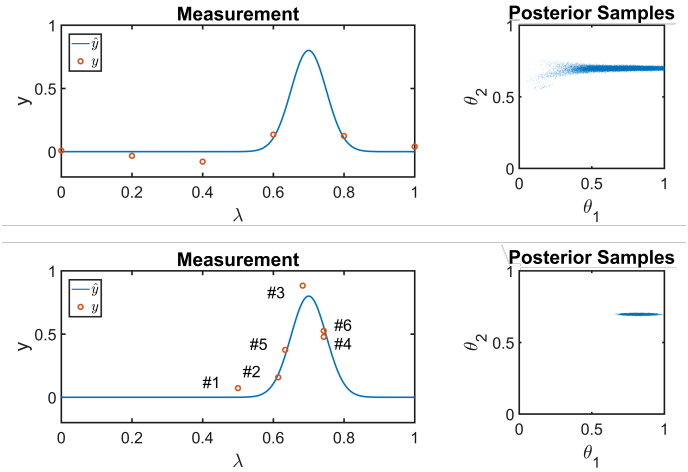

Figure 3.2: Comparison between equidistant measurement and adaptive measurement. First row: equidistant measurement and its corresponding posterior estimation. Second row: adaptive measurement and its corresponding posterior estimation.

4 Accelerated Experimental Design with RNN

As discussed in Section 3, the utility function in Bayesian experimental design can be approximated by a nested Monte Carlo method:

$$U(d) \approx \hat{U}_{N,M} \equiv \frac{1}{N}\sum_{i=1}^{N}\Big\{\log p(y^i|\boldsymbol{\theta}^i,d) - \log\big[\frac{1}{M}\sum_{j=1}^{M} p(y^i|\boldsymbol{\theta}^{ij},d)\big]\Big\}$$

where $\{\boldsymbol{\theta}^i\} \cup \{\boldsymbol{\theta}^{ij}\}$ are drawn from $p(\boldsymbol{\theta}|Y)$ and $\{y^i\}$ are drawn from $p(y|\boldsymbol{\theta}^i,d)$.

One major disadvantage of Bayesian experimental design is its slow speed. The nested MC approximation of the utility function shown above is only asymptotically unbiased as an estimator of the utility function. The bias and variance of estimator depends on the number of posterior samples. As shown in previous study [Rya03], the variance can be represented as $A(d)/N + B(d)/NM$ and the bias can be represented to leading order by $C(d)/M$, where A, B and C are terms depending on the sampling distribution. The number of samples needed for experimental design is naturally much larger than that for inference. To make things even worse, the inner loop of this nested MC is performed for each design can-

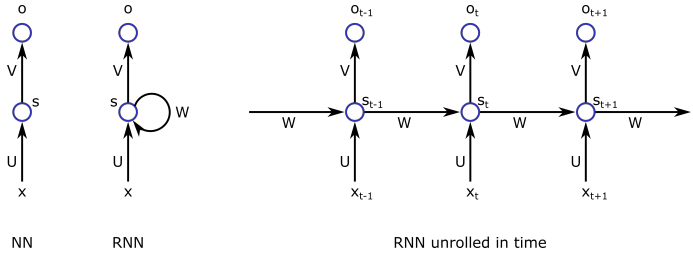

Figure 4.1: Graph representations of feed-forward neural network and recurrent neural network.

didate respectively. Due to these reasons, even with faster computers nowadays, full Bayesian experimental design is only implemented in limited fields, such as pharmaceutical studies and astronomy. What's common about these fields is that although the model behind is often very complex, the time interval between two experiments are also very long, thus allowing a good design to be found in a Bayesian way.

The problem of measuring the chromatic confocal signal is exactly opposite. Real-time decisions have to be made based on a relatively simple model. If the design speed is not fast enough, it would be more efficient to simply scan the whole wavelength range like a spectrometer. To accelerate the Bayesian experimental design process, a specific type of neural network, i.e. recurrent neural network (RNN) can be trained as an approximation.

The inspiration for using this model originates from a recent topic in computer vison society, called Visual Attention Model [BMK14]. For pattern recognition task, the researchers try to mimic the human vision system using a recurrent network. Instead of performing classification on the complete image, a small image patch is processed by the RNN and the output is both the classification result and where to look next. The training is implemented with reinforcement learning. It seems quite obvious that the visual attention model and Bayesian experimental design share an incredible amount of similarities as both attempt to gain more information through a series of adaptive measurements/observations.

For a conventional feed-forward neural network with a single hidden layer, the propagation of data can be expressed as:

$$s = f_1(Ux + b_1)$$
$$o = f_2(Vs + b_2)$$

where x denotes the input signal, o and s represent the activation of the hidden

layer and the output layer respectively. The matrices U and V contain weights describing connections between layers. The non-linear activation functions with various choices are labeled as $f_1(\cdot)$ and $f_2(\cdot)$ with b_1 and b_2 representing biases. More layers can be added to form more complex networks.

A recurrent neural network is capable of "memorizing" previous input data due to the introduction of a feedback loop in the hidden layer. Although more sophisticated variations have been developed, the simplest form of a RNN can be expressed as:

$$\begin{aligned} s_t &= f_1(Ux_t + Ws_{t-1} + b_1) \\ o_t &= f_2(Vs_t + b_2) \end{aligned}$$

where t stands for the time-stamp and W is a matrix describing the weights of the feedback loop.

To train a RNN for approximation of Bayesian experimental design, a series of experiments are simulated based on the measurement model and full Bayesian experimental design. Each simulated experiment consists of ten measurement steps of one chromatic confocal peak. The measurements and the corresponding utility functions are stored as training data for the RNN, which can be expressed in the following form:

$$\begin{aligned} l_t &= U\lambda_t + b_l \\ m_t &= Vi_t + b_m \\ s_t &= l_t \circ m_t \\ k_t &= LSTM(W, s_t, s_{t-1}, b_k) \\ o_t &= ReLU(Ok_t + b_o) \end{aligned}$$

where l_t is a hidden layer with 200 neurons to encode the measurement location, m_t is a hidden layer also with 200 neurons to encode the measured intensity. s_t merges l_t and m_t by taking element-wise multiplication with the Hadamard operator denoted by $\circ$. k_t is a sophisticated recurrent layer, namely Long Short-Term Memory (LSTM) [HS97], which memorizes information from previous measurement steps of an experiment. o_t is the output layer with rectified linear unit (ReLU) as the activation function. U, V, W, O contains weights describing each layer and b_l, b_m, b_k, b_o describe biases for the corresponding layers. The target of training is to find the weights and biases which best fit the simulated experiments and the training is conducted through RMSProp optimizer with the objective of minimizing the mean squared logarithmic error. The whole process is implemented in Python based on Theano and Keras, and is computed using

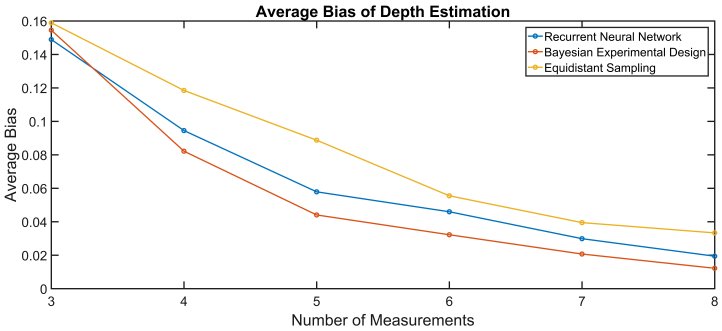

Figure 4.2: Comparison of different measurement schemes.

Quadro K1100M graphics card by Nvidia. The training takes a couple of hours, but during measurement, the feed-forward calculation of a RNN is much faster than full Bayesian experimental design which requires multiple nested MC samplings.

As a comparison, 300 experiments of chromatic confocal measurements are simulated using three approaches: full Bayesian experimental design, approximation using RNN, and equidistant measurement. Parameters of the signal are drawn randomly. As can be seen from Figure 4.2, measurement with Bayesian experimental design has a lower average estimation bias compared with an equidistant measurement method when the number of measurements are equal. The approximation by recurrent neural network does not perform as well as the Bayesian experimental design, due to the errors generated in the utility functions. However, it still yields lower bias for parameter estimation compared with the equidistant measurement scheme.

Conventional feed-forward neural network with even just a single hidden layer, is proven to be a universal approximator [Cyb89]. It means that any function can be approximated by a neural network with a single hidden layer as long as the layer is large enough. RNN is even better and has been proven to be Turing-complete [Sie95]. While training of feed-forward neural network can be seen as optimization over functions, training of recurrent neural network can be seen as optimization over programs. There exists theoretically one RNN which perfectly approximates Bayesian experimental design of a specific model.

5 Conclusion

In this article, Bayesian experimental design is applied to chromatic confocal measurement technology in order to accelerate the measurement process through an adaptive procedure. Simulations based on Markov-Chain Monte Carlo sampling are conducted to calculate the expected Kullback-Leibler divergence of the posterior distribution after a new measurement step, which serves as the utility function for guiding the next measurement step. Although experiments conducted through Bayesian experimental design demonstrate better parameter inference accuracy than common equidistant measurement schemes, the computation for the nested MC approximation of the utility function is very intensive. Therefore, a recurrent neural network based on LSTM is trained in order to approximate full Bayesian experimental design with much faster speed. Simulations have shown that accelerated experimental design with RNN provides better results than an equidistant sampling scheme and performance close to full Bayesian experimental design.

All results presented in this report are derived from simulations based on an ideal measurement model with Gaussian-shaped confocal signal. Although widely accepted, such Gaussian signal model might not truly reflect practical situations, where the confocal peak can be asymmetric depending on the specific optical setup. Therefore, a more accurate measurement model should be built based on the actual hardware implementation. Real measurements based on Bayesian experimental design should be recorded to serve as the training data for the RNN to get a more valid approximation.

Bibliography

[BMK14] J. Ba, V. Mnih, and K. Kavukcuoglu. Multiple object recognition with visual attention. *CoRR*, abs/1412.7755, 2014.

[Cyb89] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, 1989.

[DLY+13] D. Dan, M. Lei, B. Yao, W. Wang, M. Winterhalder, A. Zumbusch, Y. Qi, L. Xia, S. Yan, Y. Yang, P. Gao, T. Ye, and W. Zhao. DMD-based LED-illumination Super-resolution and optical sectioning microscopy. *Scientific Reports*, 3:1116, January 2013.

[EAH02] A. Egner, V. Andresen, and S. W. Hell. Comparison of the axial resolution of practical nipkow-disk confocal fluorescence microscopy with that of multifocal multiphoton microscopy: theory and experiment. *Journal of Microscopy*, 206(1):24–32, 2002.

[GW] J. Goodman and J. Weare. Ensemble samplers with affine invariance. *Communications in applied mathematics and computational science*, 5(1):65–80.

[HS97] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, Nov 1997.

[HWS81] D. K. Hamilton, T. Wilson, and C. J. R. Sheppard. Experimental observations of the depth-discrimination properties of scanning microscopes. *Opt. Lett.*, 6(12):625–626, Dec 1981.

[MPPQ84] G. Molesini, G. Pedrini, P. Poggi, and F. Quercioli. Focus-wavelength encoded optical profilometer. *Optics Communications*, 49(4):229 – 233, 1984.

[Rya03] K. J. Ryan. Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *Journal of Computational and Graphical Statistics*, 12(3):585–603, 2003.

[Sie95] H. T. Siegelmann. Computation beyond the turing limit. *Science*, 268(5210):545–548, 1995.