Gradient Tree Boosting-Based Positioning Method for Monolithic Scintillator Crystals in Positron Emission Tomography

Florian Müller, David Schug, Patrick Hallen, Jan Grahe, and Volkmar Schulz

***Abstract*—Monolithic crystals are considered as an alternative for complex segmented scintillator arrays in positron emission tomography systems. Monoliths provide high sensitivity, good timing, and energy resolution while being cheaper than highly segmented arrays. Furthermore, monoliths enable intrinsic depth of interaction capabilities and good spatial resolutions (SRs) mostly based on statistical calibrations. To widely translate monoliths into clinical applications, a time-efficient calibration method and a positioning algorithm implementable in system architecture such as field-programmable gate arrays (FPGAs) are required. We present a novel positioning algorithm based on gradient tree boosting (GTB) and a fast fan beam calibration requiring less than 1 h per detector block. GTB is a supervised machine learning technique building a set of sequential binary decisions (decision trees). The algorithm handles different sets of input features, their combinations and partially missing data. GTB models are strongly adaptable influencing both the positioning performance and the memory requirement of trained positioning models. For an FPGA-implementation, the memory requirement is the limiting aspect. We demonstrate a general optimization and propose two different optimization scenarios: one without compromising on positioning performance and one optimizing the positioning performance for a given memory restriction. For a 12 mm high LYSO-block, we achieve an SR better than 1.4 mm FWHM.**

***Index Terms*—Field-programmable gate array (FPGA), gradient tree boosting, machine learning, monolithic scintillator, positron emission tomography (PET).**

I. Introduction

POSITRON emission tomography (PET) is a functional imaging technique with high sensitivity manifoldly utilized in preclinical and clinical applications. To detect the two 511 keV gamma particles originating from a positron-electron annihilation, these gamma particles are converted into optical photons by scintillation crystals. The optical photons are registered by photomultiplier tubes, avalanche diodes or silicon photomultipliers (SiPM). Our group presented an MR-compatible preclinical PET insert employing segmented crystal arrays read out by digital SiPM (dSiPM) [1]. To improve the spatial resolution (SR) and to allow the determination of depth of interaction (DOI), one possibility is to reduce the pitch size and introduce multiple layers of segmented crystal elements separated by thin insensitive material significantly increasing the costs by a factor of 3–6.

Monolithic crystals are widely considered an alternative in research [2]–[7]. Monoliths are easier to fabricate than segmented arrays and have a higher sensitivity due to the reduction of insensitive material required for the segmentation. Several studies have shown good coincidence resolving times [3] and energy resolutions [2]. Furthermore, monoliths provide a SR better than 2 mm FWHM [2] and DOI can be directly derived from the light distribution [8]–[10].

However, monoliths are not yet widely employed in clinical systems due to different reasons described in the following. To achieve high SR, mostly statistical positioning models are employed. The detector is illuminated with a parallel hole collimated gamma beam at known positions and the response is measured to create reliable positioning models. A wide range of positioning algorithms including maximum likelihood (ML) estimation [11], neural networks [12], [13], support vector machines [4], and k nearest neighbor searches (kNN) [9], [14] have been proposed. However, methods based on parallel hole collimated data require calibration times of days up to weeks [9]. Such calibration times seem unlikely to be feasible when calibrating a whole PET ring. One possible solution to reduce the calibration time to hours is the utilization of a fan beam collimator [9]. Until now, this calibration method is experimentally validated only for the kNN algorithm on the level of single detector blocks [15].

The feasibility to implement the positioning algorithm in a system architecture for a large number of detector stacks is another important point to translate monoliths into clinical systems. An implementation of the positioning algorithm in an field-programmable gate array (FPGA) is advantageous to reduce the amount of data which has to be transferred out of the PET system to a control PC. The kNN algorithm is computing-intensive as a distance metric for each event under test with all training events is calculated. Assuming m training events of dimension c, the complexity is $\mathcal{O}(mc)$ for calculating the distance metrices. The found distances need to be sorted to find the kNNs requiring additionally $\mathcal{O}(mk)$ which leads to $\mathcal{O}(mc + mk)$ in total. The memory requirement is $\mathcal{O}(mc)$.

Manuscript received February 27, 2018; revised April 13, 2018; accepted May 11, 2018. Date of publication May 17, 2018; date of current version August 31, 2018. This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant 667211. *(Corresponding author: Florian Müller.)*

The authors are with the Department of Physics of Molecular Imaging Systems, Institute for Experimental Molecular Imaging, RWTH Aachen University, 52074 Aachen, Germany (e-mail: florian.mueller@pmi.rwth-aachen.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRPMS.2018.2837738

The kNN algorithm can be speeded up by prepositioning events and searching the nearest neighbors only in a subset of the reference data [2]. However, the memory requirement governed by the number of training events remain. Based on the data given in [2], we estimate a memory requirement of more than 800 MB for a single detector stack of scintillator-dimensions of 32 mm × 32 mm × 22 mm. Although it may be possible to reduce the memory requirement for scintillators of smaller height, this memory requirement is impractical for currently available FPGAs. Adding external memory would overcome this, but would add further complexity in the system design at higher costs, requires additional space and increases the power consumption. The increased power consumption and the space requirements are especially critical for highly integrated PET and PET/MR systems. Thus, it is of high interest to find computationally efficient positioning algorithms with low memory requirements.

In this paper, we present a calibration method capable to utilize both parallel hole and fan beam collimated data. Employing the fan beam collimator, a full calibration for planar positioning requires less than 1 h. We demonstrate a positioning method based on gradient tree boosting (GTB) regression. GTB is a supervised machine learning method building predictive models organized as an independently evaluable set of chains of binary decisions (decision trees). Thus, determining the position of an event is fully parallelizable and computational efficient because only simple comparisons with two possible outcomes are evaluated. The positioning performance and memory requirement of the trained models can be influenced during the training process. An FPGA implementation is already shown while the memory requirement of the models is the limiting factor to fit the available memory of the FPGA [16], [17]. We present two different optimization scenarios: one optimized for a high positioning performance and one to find the best positioning performance for given memory restrictions.

II. Materials

We used the technology evaluation kit (TEK) of Philips Digital Photon Counting (PDPC) as a coincidence setup to read out two sensor tiles built up from DPC 3200-22 digital photon counters. As exchangeable collimators, a parallel hole as well as a fan beam collimator were utilized. A monolithic LYSO crystal of 12 mm height was studied. For detecting coincidences, a pixelated array was chosen. The complete setup was placed in a light-tight temperature chamber.

A. Photodetectors

We used an array, also referred to as tile, of 16 independent dSiPM DPC 3200-22 of dimensions of 32.6 mm × 32.6 mm from PDPC [18]. Each DPC consists of four pixels resulting in a photosensor with 64 pixels and a pixel pitch of 4 mm. Each pixel contains 3200 single photon avalanche diodes (SPAD) on an active area of 3.2 mm × 3.88 mm. Every SPAD is connected to an individual logic circuit for charging and readout. We deactivated 10% of the noisiest SPADs to reduce the overall dark counts based on a dark count measurement [19]. The sensor applies a configurable two-level trigger scheme to detect and validate gamma particle interactions. Applying trigger scheme 2, 2.33 detected photons are required on average to generate a trigger signal [20]. Then, a validation interval starts within which the second threshold has to be fulfilled. On average 17 photons need to be detected (validation setting 0x55:OR) to validate a trigger and start the integration time [20]. If one pixel validates, all four pixels of the corresponding DPC are read out. The information of one DPC are referred to as hit. As stated before, every DPC is independent. Subsequently, not all 16 DPCs of the tile necessarily trigger and validate the trigger generating hits, especially for low photon densities. The tile offers a neighbor logic feature to force a read-out of the whole tile [21]. Neighbor logic is not applied to reduce dead time of the whole tile caused by inappropriate validations. More details of the sensor tile can be found in [22].

B. Scintillator Crystals and Wrappings

A monolithic LYSO crystal (Epic Crystals, Kunshan, Jiangsu, China) with a ground plane of 32 mm × 32 mm matching the active sensor tile area and 12 mm height was studied. The crystal was wrapped with highly reflective Teflon tape (Klinger, Idstein, Germany). The monolith was coupled to the tile with the two-component dielectric silicon gel Sylgard 527 (Dow Corning, Midland, MI, USA). As coincidence detector, we used a 12 mm high pixelated array with a pitch of 1 mm as employed in [23] and [24].

C. Collimator Setup

The whole setup was placed in a light-tight temperature chamber. Small fans additionally cooled the photodetectors to 5 °C. In a PET, respectively, PET/MR system, the chosen temperature can be achieved with a liquid cooling system as demonstrated for the Hyperion II^D insert [1]. The detector under study was placed on the electrically driven two-axis translation stage LIMES 90 (OWIS, Staufen im Breisgau, Germany). The translation stage sends its position by a feedback loop to a control PC. The maximum position repetition error is specified as 2 μm by the manufacturer. Up to two ^{22}Na sources with an active diameter of 0.5 mm and an activity of approximately 10 MBq each were used with both collimators. The radioactive sodium salt is backed in epoxy and encapsulated in an acrylic cylinder of 25.4 mm diameter and 6 mm height. The coincidence setup was operated with two exchangeable collimators.

1) Parallel Hole Collimator: The parallel hole collimator has a length of 51 mm and bore diameter of 0.5 mm. An additional lead shielding enclosing the collimator and sources suppresses random coincidences and scattered coincidences. A detailed description and characterization of the parallel hole collimator is given in [25].

2) Fan Beam Collimator: We present a newly developed fan beam collimator with an adaptable beam width. The working principle is based on a bottom shielding and two shielding cakes as shown in Fig. 1. The distance between shielding cakes and bottom shielding defines the beam width. To prevent a

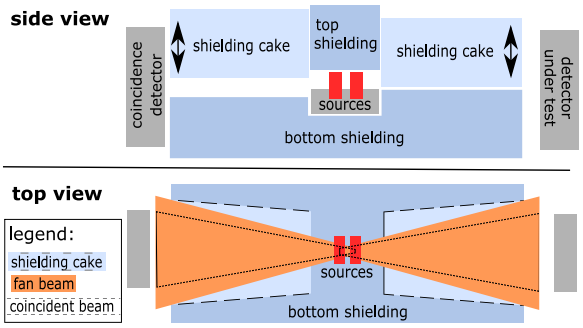

Fig. 1. Principle of the fan beam collimator. Side view: the distance between bottom shielding and shielding cakes defines the beam width. Top view: the geometrical shape of the shielding cakes restrict the span of the fan beam. The coincidence detector determines the area of the coincident beam inside the fan beam.

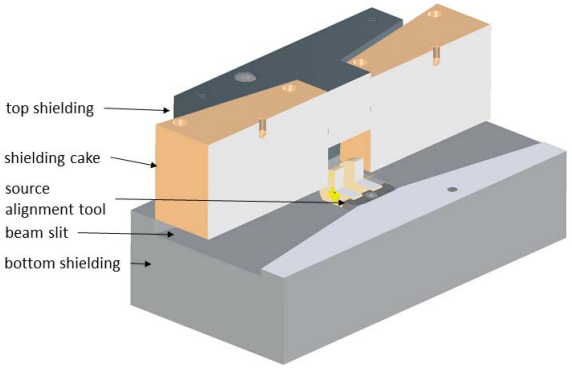

Fig. 2. Mechanical realization of the fan beam collimator as three-quarter section view. Four screws maintain the distance between bottom shielding and each shielding cake. An additional top shielding reduces scatter events and shields the sources. The source alignment tool hosts the radioactive sources inside the collimator.

loss of coincident gammas by unintended absorptions, a larger beam width was preventively chosen for the coincidence detector. The maximal span of the fan beam of 5 cm when exiting the lead block is defined by the geometrical shape of the shielding cakes.

The mechanical realization additionally contains a top shielding and a source alignment tool (see Fig. 2). The source alignment tool aligns the active area of the sources to one line parallel to the slit and changes the height of this line in the beam slit. To maintain the distance, each cake is equipped with four screws with a metric fine pitch thread of 0.5 mm/turn. The excess length of the screws can be varied between 0 mm and 14 mm. A dial indicator of 0.01 mm precision was used to determine the excess length. In order to minimize scattering, the screws are positioned outside of the direct beam path.

III. Methods

A. Beam Characterization

We determined the beam profile as described by Ritzer *et al.* [25] as well as the coincidence rate: while moving the target detector step-wise into the beam, the coincidence count rate of the setup is measured. The integral flux across the part of the beam profile already illuminating the scintillator is described by the measured coincidence rate $m(x)$ at detector position x. The beam profile $b(x)$ is the derivative of the measured rate $m(x)$. With discrete measurement points x_i, discrete derivation of the count rate $m(x)$ leads to

$$b_i := b((x_i + x_{i+1})/2) = \frac{m(x_{i+i}) - m(x_i)}{x_{i+i} - x_i}.$$

The obtained beam profile is described using a Gaussian fit. At the maximum of the beam profile, half of the beam covers the crystal. This point was assigned as the edge of the crystal. Using this method, the coordinate system of the stepper motor was aligned with a crystal coordinate system.

B. Data Acquisition and Preprocessing

All measurements were conducted at a tile temperature of 5 °C. The data acquisition process depends on the utilized collimator. In all cases, the edges of the crystal were determined by the method explained in the previous section. For the parallel hole collimator, the crystal was irradiated at defined positions on a 2-D equidistant and perpendicular grid of 0.75 mm pitch. Thus, the calibration included 1849 points over the crystal surface. Using the fan beam collimator, parallel lines of 0.25 mm pitch were irradiated. Then, the crystal was rotated by 90° and the calibration was repeated. This results in a total of 256 line measurements. For a pitch of 0.75 mm, only 86 line measurements are required.

For later analysis, the recorded data was preprocessed as described in the following employing a tool developed in this group by Schug *et al.* [23]. For both detectors, a gamma interaction generates up to 16 hits. The hits related to one gamma interaction are merged and called a cluster. The assigned timestamp of the cluster is the timestamp of the earliest hit. We used a cluster window of 40 ns to merge the hits. The TEK setup allows to apply a coincidence window on the DPC hits during the measurement. We applied a coincidence window of 40 ns for the hits in the TEK setup to account for the cluster window needed afterward. Then, coincident clusters were searched employing a sliding coincidence window of 20 ns. To distinguish pixels missing in a cluster from a zero photoncount, missing pixels are marked with a negative value. Missing pixels can be identified because the corresponding DPCs do not generate a hit and are not present while merging to clusters.

We discarded clusters with a total photoncount below a threshold of 700 photons to reject noise. Based on the obtained photon distribution, this equals an energy threshold of approximately 290 keV. On average, 10% of all clusters with the lowest photoncount were ignored. Qualified clusters are called events. No further quality cuts were applied.

The recorded data was separated in three data sets: 1) training data to train GTB models; 2) validation data to tune the hyperparameters of GTB models (validation); and 3) test data to finally determine the positioning performance of trained GTB models (evaluation). The number of events per irradiation position and the pitch of irradiation positions of the training data were varied in the following (see Section III-D1). At maximum, 10 000 events per irradiation position for the fan beam collimator and 1000 events per irradiation position for the parallel hole collimator were used. In all cases, validation and test

TABLE I
INPUT COMBINATIONS AND THEIR INCLUDED FEATURES. THE CF INCLUDE COG, MAIN DIE, MAIN PIXEL, ROW AND COLUMN SUMS

input name	input features
r	photon counts of the 64 pixels
CF	only calculated features
CF+r	CF and raw data
CF+PCA	CF and PCA
CF+PCA+r	CF, PCA and raw data

data contained events of the finest measured grid of 0.25 mm. Both validation and test data set consisted of 1250 events per irradiation position for the fan beam collimator and 125 events per irradiation position for the parallel hole collimator.

As described in more detail later on, GTB is able to handle arbitrary input features and to recognize their respective information content. Hence, we tested the influence of several input features based on the raw data of the DPCs as well as physically motivated features calculated from the raw DPC data and their combination on the positioning performance (see Table I).

In some cases and system designs, it can be beneficial to reduce the number of input features as demonstrated and discussed in [26]. We used principal component analysis (PCA) to reduce the 64 pixel counts to 16 values. On average, these 16 values represented 80% of the total information for our data. Details on PCA and the used scikit-learn implementation can be found in [27] and [28].

Statistical positioning algorithms have no information about the physical properties of the given problem. For example, the center of gravity (COG) is strongly correlated with the interaction position of the gamma particle. Using such features or adding them to the raw data can be beneficial for the positioning performance of the GTB models. We defined a set of features referred to as calculated features (CF) in the following. The CF contained index numbers of DPC and pixel with the highest photon count (main DPC and main pixel), the COG position as well as the row and column sums. It is emphasized that all CF can directly be calculated on base of the raw data, no further information are required. Missing DPC hit information in an event influences both the CFs and PCA leading to an uncertainty or jitter. The GTB needs to detect this uncertainty and to adapt it during the training process.

C. Performance Parameters

We used several parameters to test and characterize the performance of the positioning models. Using the parallel hole collimator, the performance parameters are based on the point spread function (PSF) of the detector response. The PSF is defined as the 2-D distribution of the positioning errors in both x- and y-direction. Using a fan beam collimator, only one of the spatial coordinates is known from the collimator position. Thus, performance parameters are calculated based on the line spread function (LSF) of the detector defined as the 1-D positioning error distribution. The known collimator position defines the true position of an event. The following performance parameters are employed.

1) *Bias Vector:* The bias vector represents the mean positioning error at a given position. In general, the distribution of the bias vector magnitude follows no Gaussian profile due to edge effects. Thus, we report the median and the 90th percentile of the bias vector magnitude distribution probing the central region and the tails of the distribution, respectively.
2) *SR:* The SR is defined as the full width at half maximum of the projected PSF or LSF. The fitting procedure was performed in accordance to the NEMA NU 4-2008 standard [29]. The SR is not sensitive to bias vector effects.
3) *Intrinsic Spatial Resolution (SR*):* The SR* is an estimate for the intrinsic detector resolution. We corrected the SR by quadratically subtracting the finite beam diameter determined as described in Section III-A.
4) *Percentile Radius r_x:* The percentile radius r_x is the radius enclosing the given percentile x of all assigned events around the true irradiation position. Thus, the percentile radius is sensitive to the bias vector. For percentile radii based on PSF, the Euclidean distance is employed with no projection needed. As for the bias vector distribution, we report the r_{50} and the r_{90}.
5) *Score of Radius 1.5 mm:* The score value is the fraction of correctly assigned events. An event is called correctly assigned if it is found within the given radius of 1.5 mm around the collimator position. As the percentile radius, the score value is sensitive to the bias vector.

D. Gradient Tree Boosting Position Estimation

As a supervised machine learning technique, GTB employs training data with known irradiation positions to build predictive regression models. In the following, the main aspects of the algorithm are described using a toy model of a 1-D crystal coupled to a four-channel photosensor made up of two 1-D DPCs with two channels each [see Fig. 3(a)]. Detailed reviews of decision trees and gradient boosting can be found in [30] and [31]. A mathematical description of the employed XGBoost implementation is given in [32].

Like many other machine learning algorithms, GTB tends to a phenomenon called overfitting; the residual variation as it was part of the underlying set of training data is introduced into the model, decreasing the performance of the predictions on unknown data [33]. To avoid overfitting effects, the hyperparameters of the GTB models introduced in the following were tuned with the validation data test. The final evaluation of the positioning performance was based on the test data set not invoked either in training nor in validation of the GTB models. Both validation and evaluation employed the performance parameters described before.

Trained GTB models are ensembles of so-called decision trees [see Fig. 3(b)]. The ensemble is trained in an additive manner: the first decision tree is based on the given irradiation position. Every following decision tree is trained on the residuals of the previous ensemble (irradiation position–estimated position). Thus, every newly added decision tree corrects the results of the previous ensemble. We use an additional factor

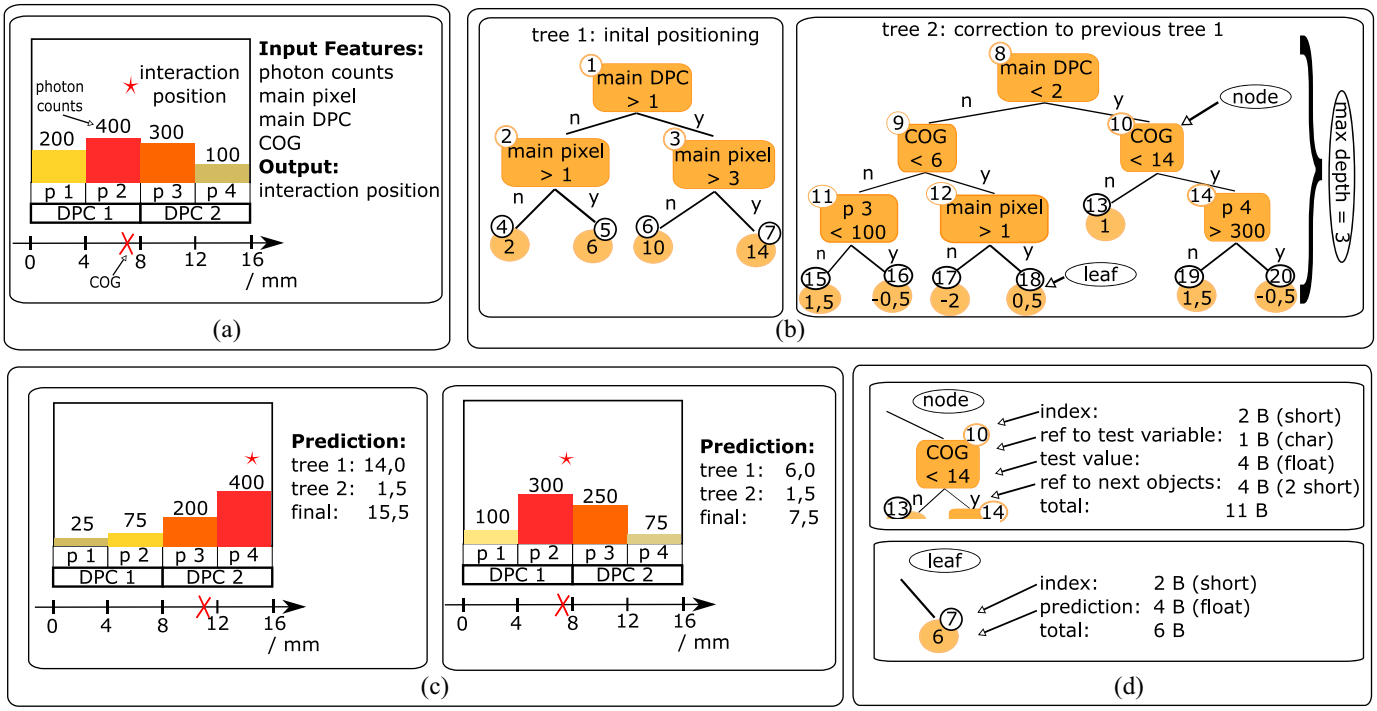

Fig. 3. Illustration of gradient tree boosting. (a) A 1-D crystal is coupled to a four-channel photosensor. Each two pixels are grouped to one DPC in this example. The listed input features are used to train a GTB model with the interaction position as output. (b) The trained GTB models contain two decision trees. Every decision tree is a sequential chain of binary decisions leading to a prediction. The second decision tree corrects the prediction of the first one. (c) Two evaluation examples are shown and the results employing the GTB model in (b). Both decision trees can be evaluated in parallel. The final prediction is the sum of all predictions of the ensemble. (d) Estimation of the memory requirement of a node and a leaf employing C standard library data types.

(< 1) called learning rate scaling the residuals and thereby the references for the next training. A smaller learning rate reduces the influence of a single tree leaving space for further optimization. Decreasing the learning rates improves the performance of trained models at the cost of needing a larger number of decision trees [34].

Every single decision tree is a sequential model combining a chain of binary decisions (node) leading to a prediction [leaf, see Fig. 3(b)]. A single node evaluates one input variable with a single split value. The maximum number of binary decisions in a single decision tree is called maximum depth d. While the training is an additive process, the final decision trees can be completely independently evaluated in parallel to find the position of an event under test. The final prediction is the sum of all predictions of the ensemble [see evaluation examples in Fig. 3(c)]. Due to the additive training based on residuals, the corrections get smaller for decision trees of higher order. In the given example, the decision trees output a prediction for the interaction position of the gamma based on the input features such as the photon counts and the COG.

Comparable to the ensemble, every single decision tree is trained in an additive manner as well. The algorithm starts with the first node and greedily adds these new nodes improving the objective loss function of current decision tree most. The root mean squared error (RMSE) is employed as objective loss function. No further nodes are added if the maximum depth is reached or the improvement is not justified compared to the higher complexity penalized by a loss function. In this way, GTB automatically chooses the features with the highest information content from the provided set of input features. Features with low information content, for example, caused by uncertainty or jitter as mentioned before, are less likely to be considered. In our example, the main DPC is used earlier than single photon counts due to the higher information content.

GTB handles missing features in both training models and prediction of events. Thus, nearly all data including partially read out clusters can be used for calibration and positioning reducing the required calibration times. In case of sensors which are self-triggering and validating on pixel level, a trigger caused by dark counts leads to a validation phase. During this validation phase, triggers of a real gamma interaction cannot be detected causing a dead time. This reduces the rate of events with information of all pixels present. Using a positioning algorithm able to handle such events opens up the potential to reduce dead-time effects and to increase the sensitivity of a system compared to algorithms requiring complete data such as standard kNN-implementations.

As motivated before, the future goal is an implementation of trained GTB in FPGAs. The feasibility of an implementation is already shown while the memory requirement of the models is the limiting factor to fit the available memory of the FPGA [16], [17]. C data types of the standard library are used to estimate the memory requirement. Further optimization might be possible when using optimized data types for an FPGA implementation. A single node object requires a total of 11 B and a leaf object 6 B [see Fig. 3(d)]. The number of nodes is less than or equal to $2^d - 1$. At maximum, 2^d leaves are present. Thus, the total memory requirement MR

of a single decision tree is

$$\mathrm{MR}(d) = \left(2^d - 1\right) \cdot 11\,B + 2^d \cdot 6\,B. \tag{1}$$

The training and output of one GTB model is strictly 1-D and completely independent for every direction. Therefore, two separate models for *x*- and *y*-direction are needed. Subsequently, both positioning steps are parallelizable in a system architecture. The separation of both positioning directions enables the use of a fan beam collimator.

The introduced hyperparameters, number of decision trees, maximum depth, learning rate, and the input features directly influence the performance and the memory requirement of GTB models. In the following, their effect is discussed and two optimization scenarios are described: 1) a high-performance optimization without considering the memory requirements and 2) an optimization process to find the best performing models fulfilling given memory-requirement restrictions.

1) General Optimization Process: We developed a protocol to study the influence of several parameters which can be divided up into three areas: 1) the measurement process (binning of the calibration grid, number of events per calibration point); 2) the algorithm hyperparameters (number of decision trees, maximum depth, learning rate); and 3) the input parameters (as defined in Table I). Trained models were verified based on all defined performance parameters with the validation data to avoid overfitting. First, we randomly tested different combinations of the given parameters to find a suitable start point for the following process. The parameter ranges for the initial search were adapted to suggestions found in [31]. The course off all performance parameters as well as the objective loss function of the GTB models were plotted against the number of decision trees to check the dependencies between the single parameters. Second, always one of the listed parameters was varied keeping all other parameters constant. For all parameters regarding the measurement process and the algorithm parameters, GTB models for up to 1000 decision trees were trained. For the different input features, GTB models of a fixed number of decision trees were trained varying the maximum depths. Based on the variation of the maximum depth, an ensemble size of 100 decision trees was selected. We chose to present the averaged 90th percentile radius as validation metric to include bias effects and to be sensitive to the tails of the LSF or PSF, respectively.

2) High-Performance Optimization: Based on the results of the general optimization process (see Section V), the binning of the calibration grid was set to 0.75 mm, the number of training events to 5000 for the fan beam collimator (250 for the parallel hole collimator) per irradiation position and the input set to CFs and raw data (CF+r). Then, we conducted an exhaustive search validating all possible combinations of the algorithm hyperparameters: The discrete hyperparameters, number of decision trees and maximum depth, were varied ranging from 1 to 1000 and from 4 to 12 by steps of 1, respectively. The continuous learning rate was probed for the values [0.05, 0.1, 0.2, 0.3, 0.4, 0.7].

To select the GTB models of the best positioning performance, we adapted the classical optimization problem of 1-D functions. The course of all performance parameters was discretely derived with respect to the number of decision trees. The first intersections with the ordinate axis were searched for every performance parameter excluding the SR (see Section V) leading to a list of numbers of decision trees. Then, starting for the highest number of decision trees, the other parameters were tested for overfitting effects. A performance parameter was assumed to be overfitting if the derivative was positive (negative for the score value). In case one parameter was overfitting, the next highest number of decision trees was tested repeating the procedure. The final evaluation of the GTB models was performed using the test data.

3) Memory-Requirement-Performance Optimization: The limiting factor for an FPGA implementation is the memory requirement of trained models as described before. Thus, it is of high interest to find the best performing models for given memory resources. Employing the same set of parameters as listed above (see Section III-D2), we trained models for different learning rates and maximum depths. In contrast to the high-performance-optimization, an empirically suitable convergence criterion based on the r_{90} determined the number of decision trees in the ensemble: the number of decision trees was fixed if the newly added decision tree improved the averaged r_{90} less than 0.01 mm. The r_{90} and the SR were plotted for all found models against the memory requirement allowing to find the best suitable models.

TABLE II
USED BEAM WIDTHS AND COINCIDENCE RATES OF PARALLEL HOLE AND FAN BEAM COLLIMATOR

parallel hole collimator		
bore diameter	beam width	coincidence rate
0.5 mm	(0.723 ± 0.022) mm	$(24 \pm 3)\ \mathrm{s}^{-1}$
fan beam collimator		
slit width	beam width	coincidence rate
0.25mm	(0.417 ± 0.004) mm	$(199 \pm 9)\ \mathrm{s}^{-1}$

IV. RESULTS

A. Beam Characterization

The results of the beam characterization are summarized in Table II. The parallel hole collimator reached a coincidence rate of $24\mathrm{s}^{-1}$ and the fan beam collimator of $199\mathrm{s}^{-1}$ for a significant narrower beam width which equals a factor of about 8.

B. Data Acquisition

Taking all irradiation positions into account, the photopeak position calculated based on the total photonsum was found for 1240 photons with an FWHM of 396 photons which equals an energy resolution of approximately 32 %. Fig. 4 shows obtained read-out characteristics with a Gaussian fit. The Gaussian distribution describes the data well.

C. General Optimization Process

The randomized parameter search lead to a maximum depth of 10, a learning rate of 0.1 and raw data of 5000 training

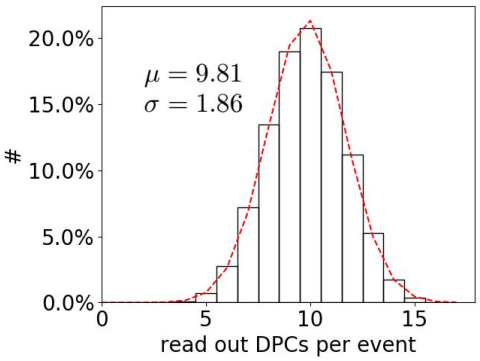

Fig. 4. Histogram and Gaussian fit (dashed line) of read out DPCs for events acquired by a homogeneous illumination of the crystal surface.

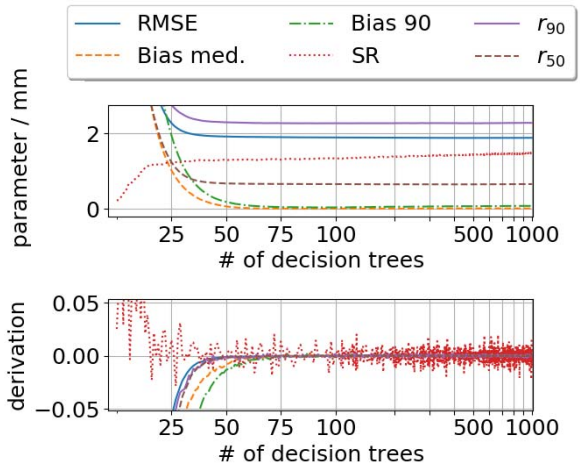

Fig. 5. Performance parameters and the objective loss function (RMSE) of the GTB models (top) and their discrete derivative (down) against the number of decision trees. The abscissa is plotted linear up to 100 decision trees and logarithmic afterward. A maximum depth of 10 and a learning rate of 0.1 were applied with raw data as input. The course of all shown performance parameters is strongly correlated. The SR increases with a higher number of decision trees. The derivative of the SR has the highest fluctuation of all performance parameters.

events acquired every 0.75 mm as start point for the training process. The GTB objective loss function RSME is strongly correlated with all performance parameters (see Fig. 5). All performance parameters except the SR continuously decrease until their optimum is reached. Their derivatives tends to 0 with a decreasing slope for a higher number of decision trees. After their optimum, the performance parameters fluctuate or worsen again. The SR increases adding more decision trees especially up to about 25 decision trees. The derivative of the SR is neither continuous nor monotonous.

Results of the general optimization process of the fan beam calibration along one direction are exemplarily shown in Fig. 6. We used the found start point for the training process if not stated otherwise.

The GTB models show a very similar course of the r_{90} for all studied pitches ranging from 0.25 mm to 0.75 mm of the calibration grid [see Fig. 6(a)]. The performance increases less than 2% for those models trained with the finest available irradiation pitch compared to the coarsest one evaluated at their respective optima. Fig. 6(b) displays the r_{90} as a function of the irradiation position for the coarse and the fine calibration grid. In the upper plot, both distributions are showing a very similar course. Their characteristics are the same in general including the behavior at the edges. A significant improvement of the r_{90} is observed close to 3 mm and 29 mm for both pitches of the calibration grid. The lower plot displays the difference between the corresponding r_{90} distributions fluctuating unsystematically around 0.

In case the number of training data is increased, the model performance improves until a sufficient amount of data is given [see Fig. 6(c)]. After this point, just small improvements can be observed if further training data are added. As an example, doubling the training events from 5000 to 10 000 per calibration point boosts the r_{90} by less than 1.5% evaluated at their respective optima.

The positioning performance of the models increases for larger maximum depths [see Fig. 6(d)]. Increasing step-wise the maximum depth, the performance-boost gets less toward a maximum depth of 10. For even higher maximum depths, no further gain of the positioning performance is observed. The positioning performance of the GTB models converges to a global optimum point for a large number of decision trees as visible toward 1000 decision tress.

For higher learning rates, the best positioning performance of the respective GTB models is reached for a smaller number of decision trees [see Fig. 6(e)]. However, the positioning performance deteriorates compared with the models employing a smaller learning rate. This is clearly visible for the learning rate of 0.7. Less than 10 decision trees are needed to reach the optimum point. However, the maximum performance is reduced about 20% compared to a learning rate of 0.05 evaluated at their respective optima.

In general, the performance increases for all evaluated combinations of input features except CF increasing the maximum depth up to 10 [see Fig. 6(f)]. For a depth of 12, only raw data as input lead to a slightly better performance. The other input combinations do not improve or slightly deteriorate. For maximum depths up to 6, all input combinations including CF lead to better results than raw data. For maximum depths larger than 6, models trained with raw data increase their performance compared to CF, PCA and combination CF+PCA. Combination CF+r and combination CF+PCA+r have almost the same characteristics.

D. High-Performance Optimization

The best positioning performance was found for GTB models of maximum depth 10, learning rate 0.1 and 71 to 77 decision trees requiring in total around 2 MB of memory for both planar directions. 2-D spatial distributions of the bias vector and r_{50} are shown for the parallel hole calibration in Fig. 7. The bias vector is randomly distributed in the central region of the crystal. At the edges, a bias toward the center is observed. The r_{50} is homogeneously distributed in the central region and deteriorates toward the crystal edges.

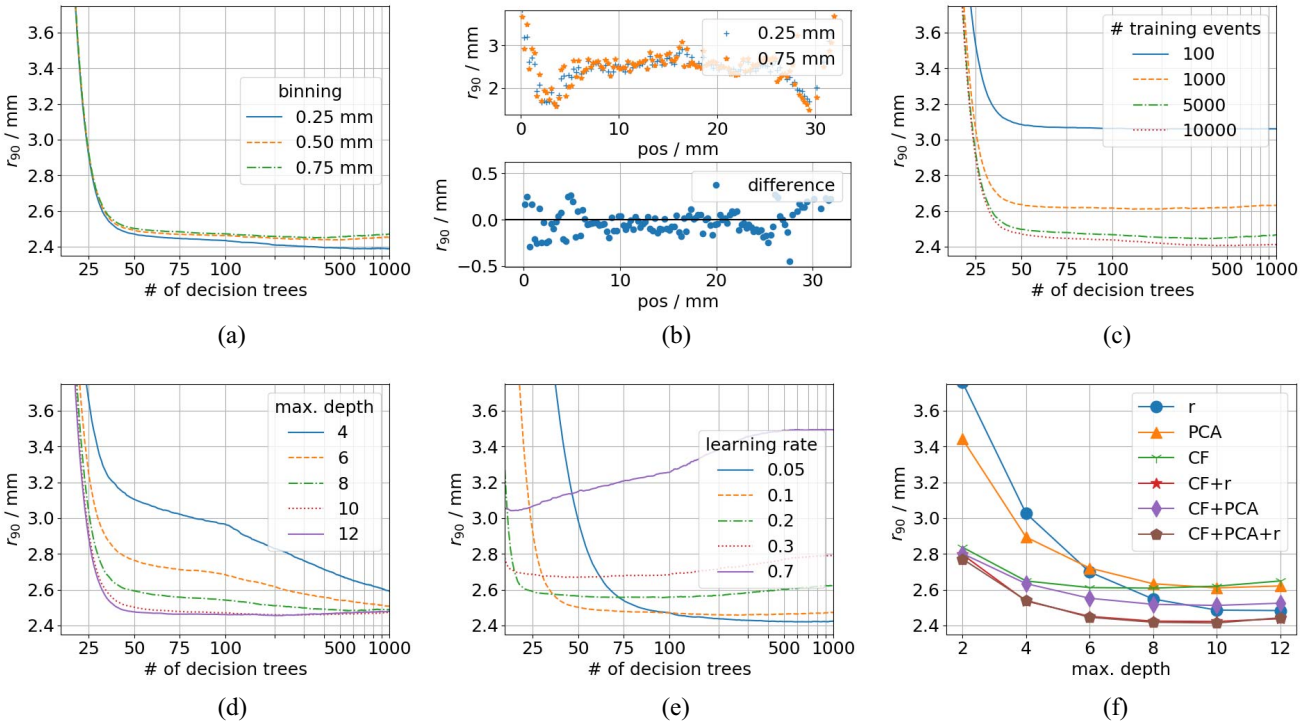

Fig. 6. Exemplary optimization process for data measured with the fan beam collimator. Unless stated otherwise, ensembles of maximum depth 10 and learning rate 0.1 are trained on raw data of 5000 training events per position on a calibration grid of 0.75 mm pitch. The test data have a pitch of 0.25 mm. In case the number of decision trees is shown on the abscissa, a linear scale is chosen up to 100 decision trees and a logarithmic scale afterward. (a) Averaged r_{90} against the number of decision trees varied for the pitch of the calibration grid. (b) Spatial r_{90}-distributions for models trained with calibration grids of 0.25 mm and 0.75 mm, respectively. Top: both r_{90}-distributions as overlay. Down: difference of the distributions. (c) Averaged r_{90} against the number of decision trees varied for the number of training events per position. (d) Averaged r_{90} against number of decision trees for different maximum depths. (e) Averaged r_{90} against number of decision trees for different learning rates. (f) Averaged r_{90} against the maximum depth for different input combinations (see Table I). All models consist of 70 decision trees.

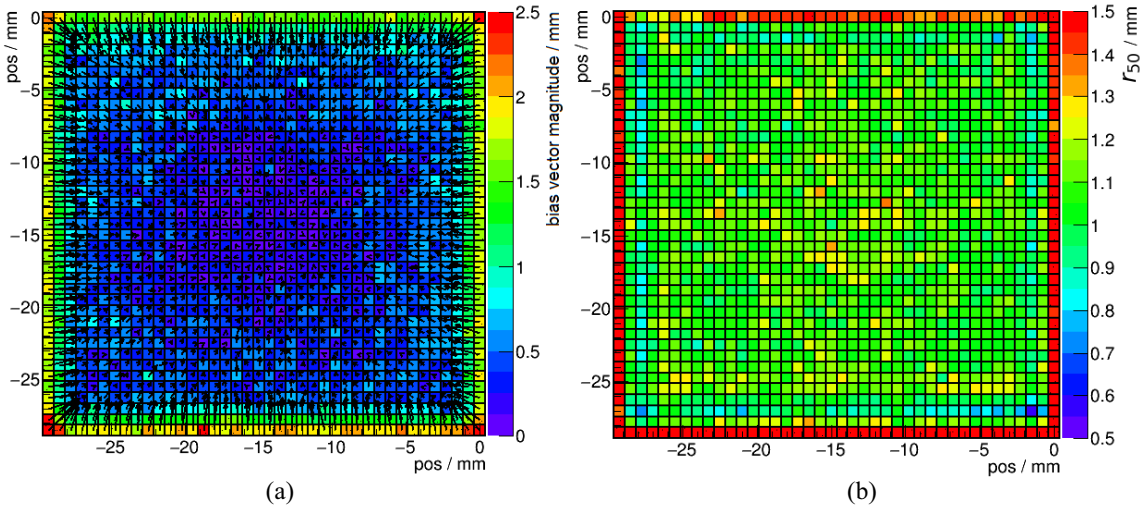

Fig. 7. Exemplary spatial distribution of the parallel hole collimator calibration based on the GTB models of the best positioning performance. (a) Bias vector distribution. The color scale represents the bias vector magnitude. (b) r_{50} distribution.

Table III shows the performance parameters for calibrations of both parallel hole and fan beam collimator. In general, the performance parameters of both calibration methods probing the central area of the LSF lead to very similar results. The r_{90} is around 6% better for the fan beam calibration. For both calibrations method, an SR* of 1.40 mm FWHM or better is achieved.

E. Memory-Requirement-Performance Optimization

Fig. 8 exemplarily shows the SR and r_{90} for possible GTB models against the memory requirement for the calibration in y-direction of the crystal. Every chosen maximum depth (denoted next to each line) is probed for several learning rates resulting in a specific number of decision trees defined by the convergence criterion and plotted as one line. For more clarity,

TABLE III
OVERVIEW OF THE BEST ACHIEVED POSITIONING PERFORMANCE FOR PARALLEL HOLE AND FAN BEAM COLLIMATED CALIBRATION. FOR THE PARALLEL HOLE CALIBRATION, PERFORMANCE PARAMETERS BASED ON THE PSF ARE GIVEN AS TOT

	parallel hole			fan beam	
	x	y	tot	x	y
SR / mm	1.55	1.56		1.46	1.31
SR* / mm	1.37	1.38		1.40	1.24
r_{50}/mm	0.66	0.67	1.14	0.66	0.65
r_{90}/mm	2.44	2.44	3.73	2.29	2.27
Bias med. / mm	0.28	0.24	0.65	0.2	0.26
Bias 90 / mm	0.74	0.88	1.55	0.84	0.85
Score 1.5	0.8	0.79	0.64	0.8	0.81

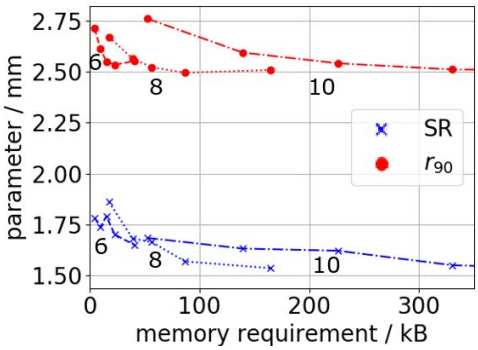

Fig. 8. Planar positioning performance calibrated in the *y*-direction of the crystal against memory requirement. The numbers next to the lines denote the maximum depth of the models.

the results of selected maximum depths are shown. For a given amount of memory, multiple combinations of maximum depth and learning rate fulfill possible memory restrictions. For maximum depths of 6 and 8, the r_{90} deteriorates while the SR still improves. GTB models of maximum depth 8 obtain a similar positioning performance compared to GTB models of maximum depth 10 with significantly less memory requirements.

V. DISCUSSION

The used method for beam characterization works well and allows a reliable determination of the crystal edges. The beam widths are larger than the bore diameter and the slit width, respectively. The beam spreads as a function of the distance between collimator and detector caused by geometrical effects and Compton-scatter of gamma particles with the collimator.

The given characteristics of the photon distribution indicate that no photopeak events are excluded for training and testing the GTB models. Due to missing hit information in events, the total photonsum is no stable energy criterion. Thus, the obtained photon distribution is not suitable for an energy calibration leading to good energy resolutions. A dedicated energy calibration needs to account for the missing hit information. Furthermore, the scintillator should be divided up into voxels with their own energy calibration to include spatial variations in the characteristics of the detector. Currently, we are investigating a dedicated energy calibration which is not in the scope of this paper.

The photon densities in the employed monolith are too low to generate a significant amount of events with hits of all 16 DPCs present (see Fig. 4). In only 0.038% of all cases, the events include hits of all 16 DPCs. This emphasizes the need for a calibration and positioning algorithm that is able to handle missing data.

The general optimization process shows the connections between the performance parameters and the influence of the algorithm parameters. The course of the SR is not as stable as these of the other defined performance parameters (see Fig. 5). As the determination procedure of the SR includes a projection and a binning process, the SR can be affected by binning artifacts. Thus, the SR is not suitable for determination of the GTB models with the best positioning performance. In the region of a small number of decision trees (up to about 25 decision trees in the shown example), the SR has very small values and worsens for increasing model ensembles. At the beginning of the training process, the possible predictions of GTB models are limited to the number of leafs. Subsequently, the predictions are discrete which is beneficial for the SR. Increasing the number of decision trees, more predictions are possible increasing the general positioning performance of the model while deteriorating the SR.

Fig. 6(a) and (b) demonstrates that the GTB algorithm builds reliable regression models. Thus, the number of calibration points and the calibration time needed can be reduced without compromising too much on the positioning performance. Therefore, a pitch of 0.75 mm of the calibration grid is employed for the training data. Most other calibration methods found in literature employ a calibration grid of 0.25 mm [2], [4], [9], [15]. Thus, GTB allows to reduce the calibration time by a factor of 3 for the fan beam collimator keeping the number of training events per irradiation position constant. The course of the r_{90} toward the scintillator surfaces is based on two effects: very close to the surfaces, the bias vector dominates the r_{90} and leads to a deterioration. Near the positions of 3 mm and 29 mm, the bias vector has vanished while the photon pattern is still affected by the reflections of the scintillator surfaces. These reflections have a higher flux compared to reflections at the surfaces caused by gamma interactions at central irradiation positions due to geometrical aspects. Thus, the reflections of gamma interactions close to the surfaces are not part of a uniform background. This leads to more distinct photon patterns compared to central positions of the crystal which is beneficial for the positioning performance of GTB models. Furthermore, the maximum width of the r_{90} distribution is geometrically limited by the scintillator surfaces. This effect is also beneficial for the r_{90}.

Beside the number of irradiation positions, the number of training events is directly proportional to the measurement time. As shown in Fig. 6(c), no significant increase in the positioning performance for more than 5000 training events is observed for the fan beam calibration. The GTB algorithm finds all causal connections in the data within this data set size. We set the number of training events per irradiation position to 5000 for the fan beam calibration and to 250 for the parallel hole calibration. This leads to a calibration time of less than 1 h for both planar directions for the fan beam calibration and

1 d for the parallel hole collimator. The GTB algorithm offers the possibility to further reduce the measurement time while still providing reliable positioning models.

Maximum depths larger than 10 are not reasonable for the presented geometry as they do not lead to a further improved accuracy [see Fig. 6(d)]. For an FPGA implementation, the maximum depth is the most critical parameter because the memory requirement is $\mathcal{O}(n2^d)$ with d the maximum depth and n the number of decision trees.

A high learning rate can be used to create well-performing positioning models with a low memory requirement. GTB models with a large learning rate tend faster to overfitting effects and perform worse compared to those models trained with small learning rates.

All models containing CF as input features show a significant increased performance compared to input combinations without them. The GTB algorithm efficiently utilizes the physical information provided by the chosen set of CF and also accounts for the uncertainties caused by missing hit information. Especially for models of a small maximum depth, the CF are easier to interpret. For higher maximum depths, the GTB model has "learned" causal connections between the raw data and is able to outperform the CF. However, adding CF to the raw data as input still improves the performance due to additional information content. Input combinations PCA and PCA+r perform worse than those input combinations containing raw data due to the information loss of the PCA. As mentioned before, 80% of the information are preserved in the first 16 PCA components. However, this demonstrates that the GTB models are able to handle PCA transformed input features. Input combinations CF+PCA+r shows no performance boost compared to CF+r because the PCA does not add additional information to the training process.

GTB models show a homogeneous positioning performance over the whole central crystal region with the typical bias effects at the edges (see Fig. 7) found in monolithic scintillators. The r_{50} deteriorates at the crystal edges because this performance parameter is sensitive to bias vector effects.

Calibrations based on parallel hole and fan beam instrumentations work well and lead to very similar results. This is due to the fact that the used GTB models are only 1-D. A 2-D implementation might benefit from point data in the sense of a more efficient memory requirement. Small deviations in single performance parameters such as the r_{90} can origin from the influence of the different size of the gamma beams. Furthermore, the rate of scattered gamma particles in the collimator may differ for both instrumentations. Considering these aspects, both calibration processes lead to nearly equivalent performances.

Comparing the positioning performance in x- and y-direction (see Table III), the GTB models perform equivalently along both directions. Small deviations may occur due to statistical effects.

Fig. 8 helps to select the best training and model parameters for a given memory restriction. Taking both displayed performance parameters into account, an optimum point for every maximum depth can be chosen before overfitting effects occur. This optimization scenario demonstrates the potential of the adaptability of the GTB algorithm. The memory requirement can be adjusted by modifying the discussed parameters. The memory requirement of the models is orders of magnitude lower compared to our estimates of 4 MB of an ML or 800 MB of a kNN implementation based on data of [2] and [11]. For example, employing the shown model of maximum depth 6 and 16 kB memory requirement (32 kB for both directions), the SR is around 1.7 mm which outperforms most pixelated clinical detector blocks used in whole-body PET [35].

VI. Conclusion

The presented GTB-based positioning algorithm allows a time-efficient calibration and is able to create positioning models suitable to be implemented on an FPGA. Compared to a parallel hole collimator, the developed fan beam collimator accelerates the full planar positioning calibration by a factor of 20 to less than 1 h. Our developed positioning algorithm flexibly handles different input features and their combinations including PCA transformed data. Calibrations with parallel hole and fan beam collimator lead to equivalent results. GTB accepts missing features for training and prediction which is beneficial for sensitivity in PET systems. Additional features based on physical properties such as the first moment (COG) significantly improve the positioning performance. The flexibility of handling all kind of input features enables possibly future optimizations. We trained GTB models for two scenarios demonstrating the versatility of the algorithm: one without compromising on positioning performance and one optimizing the positioning performance for a given memory restriction. For a 12 mm high monolithic block, we achieved an SR of 1.24 mm FHWM and 1.40 mm FWHM for y- and x-direction corrected for the finite beam size. Future work will evaluate the applicability and performance of the algorithm to DOI positioning as well as for different scintillator geometries. Furthermore, the aim of an FPGA implementation will be realized.

References

[1] B. Weissler *et al.*, "A digital preclinical PET/MRI insert and initial results," *IEEE Trans. Med. Imag.*, vol. 34, no. 11, pp. 2258–2270, Nov. 2015, doi: 10.1109/TMI.2015.2427993.

[2] G. Borghi, V. Tabacchini, and D. R. Schaart, "Towards monolithic scintillator based TOF-PET systems: Practical methods for detector calibration and operation," *Phys. Med. Biol.*, vol. 61, no. 13, pp. 4904–4928, Jul. 2016, doi: 10.1088/0031-9155/61/13/4904.

[3] H. T. Van Dam, G. Borghi, S. Seifert, and D. R. Schaart, "Sub-200 ps CRT in monolithic scintillator PET detectors using digital SiPM arrays and maximum likelihood interaction time estimation," *Phys. Med. Biol.*, vol. 58, no. 10, pp. 3243–3257, 2013, doi: 10.1088/0031-9155/58/10/3243.

[4] P. Bruyndonckx *et al.*, "Evaluation of machine learning algorithms for localization of photons in undivided scintillator blocks for PET detectors," *IEEE Trans. Nucl. Sci.*, vol. 55, no. 3, pp. 918–924, Jun. 2008, doi: 10.1109/TNS.2008.922811.

[5] R. Marcinkowski, P. Mollet, R. Van Holen, and S. Vandenberghe, "Sub-millimetre DOI detector based on monolithic LYSO and digital SiPM for a dedicated small-animal PET system," *Phys. Med. Biol.*, vol. 61, no. 5, pp. 2196–2212, 2016, doi: 10.1088/0031-9155/61/5/2196.

[6] A. J. Gonzalez *et al.*, "The MINDView brain PET detector, feasibility study based on SiPM arrays," *Nucl. Instrum. Methods Phys. Res. Section A Accelerators Spectrometers Detectors Assoc. Equipment*, vol. 818, pp. 82–90, May 2016, doi: 10.1016/j.nima.2016.02.046.

[7] A. González-Montoro *et al.*, "Detector block performance based on a monolithic LYSO crystal using a novel signal multiplexing method," *Nucl. Instrum. Methods Phys. Res. Section A Accelerators Spectrometers Detectors Assoc. Equipment*, Feb. 2018, doi: 10.1016/j.nima.2017.10.098.

[8] X. Li, C. Lockhart, T. K. Lewellen, and R. S. Miyaoka, "A high resolution, monolithic crystal, PET/MRI detector with DOI positioning capability," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2008. 2008, pp. 2287–2290, doi: 10.1109/IEMBS.2008.4649654.

[9] H. T. Van Dam *et al.*, "Improved nearest neighbor methods for gamma photon interaction position determination in monolithic scintillator PET detectors," *IEEE Trans. Nucl. Sci.*, vol. 58, no. 5, pp. 2139–2147, Oct. 2011, doi: 10.1109/TNS.2011.2150762.

[10] A. González-Montoro *et al.*, "Performance study of a large monolithic LYSO PET detector with accurate photon DOI using retroreflector layers," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 1, no. 3, pp. 229–237, May 2017, doi: 10.1109/TRPMS.2017.2692819.

[11] S. España, R. Marcinkowski, V. Keereman, S. Vandenberghe, and R. Van Holen, "DigiPET: Sub-millimeter spatial resolution small-animal PET imaging using thin monolithic scintillators," *Phys. Med. Biol.*, vol. 59, no. 13, pp. 3405–3420, 2014, doi: 10.1088/0031-9155/59/13/3405.

[12] P. Conde *et al.*, "Determination of the interaction position of gamma photons in monolithic scintillators using neural network fitting," *IEEE Trans. Nucl. Sci.*, vol. 63, no. 1, pp. 30–36, Feb. 2016, doi: 10.1109/TNS.2016.2515163.

[13] Y. Wang, W. Zhu, X. Cheng, and D. Li, "3D position estimation using an artificial neural network for a continuous scintillator PET detector," *Phys. Med. Biol.*, vol. 58, no. 5, pp. 1375–1390, Mar. 2013, doi: 10.1088/0031-9155/58/5/1375.

[14] G. Borghi, B. J. Peet, V. Tabacchini, and D. R. Schaart, "A 32 mm x 32 mm x 22 mm monolithic LYSO: Ce detector with dual-sided digital photon counter readout for ultrahigh-performance TOF-PET and TOF-PET/MRI," *Phys. Med. Biol.*, vol. 61, no. 13, pp. 4929–4949, Jun. 2016, doi: 10.1088/0031-9155/61/13/4929.

[15] G. Borghi, V. Tabacchini, S. Seifert, and D. R. Schaart, "Experimental validation of an efficient fan-beam calibration procedure for *K*-nearest neighbor position estimation in monolithic scintillator detectors," *IEEE Trans. Nucl. Sci.*, vol. 62, no. 1, pp. 57–67, Feb. 2015, doi: 10.1109/TNS.2014.2375557.

[16] R. Kułaga and M. Gorgoń, "FPGA implementation of decision trees and tree ensembles for character recognition in vivado HLS," *Image Process. Commun.*, vol. 19, nos. 2–3, pp. 71–82, 2014, doi: 10.1515/ipc-2015-0012.

[17] B. Van Essen, C. Macaraeg, M. Gokhale, and R. Prenger, "Accelerating a random forest classifier: Multi-core, GP-GPU, or FPGA?" in *Proc. IEEE 20th Int. Symp. Field Program. Custom Comput. Mach.*, IEEE, Apr. 2012, pp. 232–239, doi: 10.1109/FCCM.2012.47.

[18] C. Degenhardt *et al.*, "The digital silicon photomultiplier—A novel sensor for the detection of scintillation light," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, 2009, pp. 2383–2386, doi: 10.1109/NSSMIC.2009.5402190.

[19] T. Frach, G. Prescher, C. Degenhardt, and B. Zwaans, "The digital silicon photomultiplier—System architecture and performance evaluation," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, IEEE, Oct. 2010, pp. 1722–1727, doi: 10.1109/NSSMIC.2010.5874069.

[20] V. Tabacchini, V. Westerwoudt, G. Borghi, S. Seifert, and D. R. Schaart, "Probabilities of triggering and validation in a digital silicon photomultiplier," *J. Instrum.*, vol. 9, no. 6, pp. P06016–P06016, Jun. 2014, doi: 10.1088/1748-0221/9/06/P06016.

[21] D. Schug *et al.*, "First evaluations of the neighbor logic of the digital SiPM tile," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, IEEE, Oct. 2012, pp. 2817–2819, doi: 10.1109/NSSMIC.2012.6551642.

[22] R. Marcinkowski, S. Espana, R. Van Holen, and S. Vandenberghe, "Effects of dark counts on digital silicon photomultipliers performance," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, IEEE, Oct. 2013, pp. 1–6, doi: 10.1109/NSSMIC.2013.6829323.

[23] D. Schug *et al.*, "Data processing for a high resolution preclinical PET detector based on philips DPC digital SiPMs," *IEEE Trans. Nucl. Sci.*, vol. 62, no. 3, pp. 669–678, Jun. 2015, doi: 10.1109/TNS.2015.2420578.

[24] N. Gross-Weege, D. Schug, P. Hallen, and V. Schulz, "Maximum likelihood positioning algorithm for high-resolution PET scanners," *Med. Phys.*, vol. 43, no. 6, pp. 3049–3061, 2016, doi: 10.1118/1.4950719.

[25] C. Ritzer, P. Hallen, D. Schug, and V. Schulz, "Intercrystal scatter rejection for pixelated PET detectors," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 1, no. 2, pp. 191–200, Mar. 2017, doi: 10.1109/TNS.2017.2664921.

[26] L. A. Pierce *et al.*, "Multiplexing strategies for monolithic crystal PET detector modules," *Phys. Med. Biol.*, vol. 59, no. 18, pp. 5347–5360, Sep. 2014, doi: 10.1088/0031-9155/59/18/5347.

[27] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London Edinburgh Dublin Philos. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, 1901.

[28] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[29] *Performance Measurements of Small Animal Positron Emission Tomographs*, NEMA Standard NU 4-2008, 2008.

[30] S. B. Kotsiantis, "Decision trees: A recent overview," *Artif. Intell. Rev.*, vol. 39, no. 4, pp. 261–283, 2013, doi: 10.1007/s10462-011-9272-4.

[31] A. Natekin, A. Knoll, M.-O. Gewaltig, and O. Michel, "Gradient boosting machines, a tutorial," *Front. Neurorobot.*, vol. 7, p. 21, Dec. 2013, doi: 10.3389/fnbot.2013.00021.

[32] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Disc. Data Min. (KDD)*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[33] T. Dietterich, "Overfitting and undercomputing in machine learning," *ACM Comput. Surveys*, vol. 27, no. 3, pp. 326–327, 1995.

[34] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[35] G. B. Saha, *Basics of PET Imaging: Physics, Chemistry, and Regulations*. Cham, Switzerland: Springer Int., 2015, doi: 10.1007/978-3-319-16423-6.