euspen's 22nd International Conference & Exhibition, Geneva, CH, May/June 2022
www.euspen.eu

Machine Learning approach to predict the position of a new online position measurement system's target

Luis Garcia[1], Uwe Bielke[2], Theo Hageney[3], Klaus Banzhaf[3], Ernst Wiedenmann[4], Rainer Börret[1]

[1]Aalen University, Aalen, Germany
[2]Glasgow Caledonian University, Glasgow, UK
[3]eµmetron GmbH, Aalen, Germany
[4]Serious Enterprises, Aalen Germany

Luis.GarciaBarth@hs-aalen.de

Abstract

This article discusses a Machine Learning approach for position determination of a target, that is used in a new measurement device called Multi-Aperture-Positioning-System (MAPS). The MAPS consists of two components, the measurement unit itself and the LED-Target, which is mounted to a machine under test – for example, an industrial robot or a coordinate measurement machine (CMM). The measurement system is based on a photogrammetric approach using an aperture array and a single camera. To achieve high accuracy in position calculation, multiple complex algorithms with high computing effort are used. Actual the MAPS offers an accuracy equal to or better compared to existing photogrammetric devices.

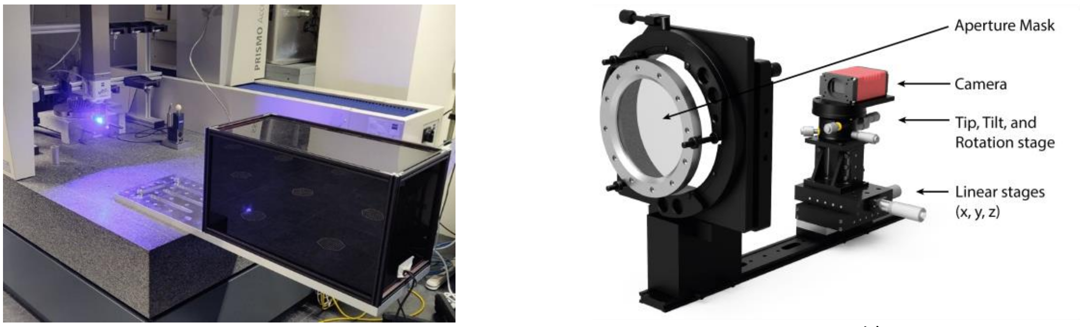
Using Machine Learning, a Convolutional Neural Network (CNN) is trained by artificial data and tested on real data. In order to generate these data sets, proprietary simulation software is used to generate artificial images. The CMM is used to move the LED-Target to different positions while MAPS takes images at these. The simulated images are used as input for training, while the real images are used to measure the performance of the CNN. Previously MAPS used several algorithms to calculate the position of the target from the taken image. The key idea behind the Machine Learning approach is, that a Neural Network, trained on thousands of labelled data, should be as accurate as the originally used algorithms in determining the position of the target while being much faster. In addition, systematic errors, model errors, and uncertainties in the system parameters can be eliminated. While the Neural Network learns directly from the captured data, algorithms use formulas to calculate the position from the images. These formulas have a lot of influencing factors, that cannot always be precisely determined, change over time, or are not even known. When using a Neural Network instead, the model learns the relation between input and output data including all influences that exist at the training time. On the whole, this paper investigates if a Neural Network can replace the originally used algorithms while achieving similar or better performance.

Keywords: Photogrammetry, 3DoF Measurement, Machine Learning, Accuracy, Online-Calibration, Neural Network

1. Introduction

Machine tools, robots and measuring machines can nowadays be found in almost every large industry like automotive, electronic manufacturing or production of components and consumer goods. As the demand for individualized product rise and the production batches shrink, multi-purpose machines become more and more important [1]. The ability to manufacture accurate parts is one of the main performance criteria for modern production. Therefore, measuring machines gain more importance as the main dimensional quality inspection instrument for manufactured parts. Particularly where high repeatability is required, measuring machines with high capability are essential. Demanding higher product quality, improving the measurement accuracy has become an extremely important area of investigation [2]. Increasing the accuracy by correcting the machines can be done in two ways. Either by correcting them in a feedback loop [3] or by calibration, applying a correction matrix [4].
Another important point in modern production is the coordination of multiple machines working on a single task or workpiece. Synchronising them is the crucial part as otherwise, they could fail. To avoid this, all machines and workpieces in the process need to be measured precisely in the same coordinate system. Measuring machines available today, such as the Laser Tracker or camera-based systems do not meet all the following criteria together: online measuring, high accuracy, and measuring multiple coordinate systems simultaneously. Therefore MAPS, a new photogrammetric measurement system introduced by Bielke et al. [5]was developed. It's made of a camera sensor, an aperture mask, and a LED-Target.

This article focuses on developing a new method to determine the centre of gravity in the light spots, which are used by MAPS to calculate the position of the LED-Target. The goal is to replace the traditionally used mathematical algorithms with an image regression Convolutional Neural Network. They have already proven their value in a wide range of applications. From vehicle detection and counting [6] to quantifying cyanobacteria [7] or the prediction of pediatric bone age assessment [8]. Neural Networks are extremely versatile since they use function approximation to best map examples of inputs to examples of outputs [9].



a) b)

Figure 1: MAPS experimental setup. a) Image of the setup mounted on a Zeiss PRISMO Access. The LED-Target is attached to a Zeiss VAST XT measurement head. The black box contains the MAPS setup which protects it from stray light and dust. b) Sketch of the MAPS setup inside the black box displayed a). The camera is mounted on a Tip, Tilt and Rotation stage on three linear axes for a movement in 6DOF. It allows the adjustment of the camera sensor to the aperture mask.

Beginning with the experimental set-up, the paper shows the components that were used in this work. Followed by the methodology, the necessary steps to calculate the position of the target with MAPS are presented. The simulation of the light spots and the training of the Neural Network by them is the major part. Explaining how the CNN is implemented is included. Applying the CNN to predict the centre of the light spots in an experiment demonstrates its performance, which is then compared to the traditionally used algorithms.

2. Experimental set-up

The experimental set-up is shown in Error! Reference source not found.. It has three major components – the measurement system MAPS, a CMM (ZEISS PRISMO Access), and the proprietary software for simulation and CNN engineering. MAPS itself consists of a high-resolution camera, an Allied Vision Prosilica GT 3300 with 8.1 megapixels which operate at around 4 fps in our set-up. In front of the camera is an aperture mask mounted. It's a photomask, coated with a layer of chromium including around 40k aperture holes. The third component is the LED-Target, amount with an ultra-bright blue LED. MAPS is mounted on the CMM's granite table facing the probe head which is the LED-Target in this set-up.

The CMM is used as a reference system for MAPS. As it moves the LED-Target to predefined coordinates while the MAPS measures the position there, the accuracy of the CMM defines the accuracy of MAPS. The images taken by MAPS are saved and later used to train the CNN. In addition to the real measurements, the simulation software is capable of simulating these images, which is explained below.

3. Methodology

To determine the position of the light source, MAPS uses a sequence of different algorithms on the taken images. An image represents a matrix of light spots, mapped by the light of the LED-Target passing through the holes of the aperture mask and hitting the camera sensor. To calculate the position of the LED-Target from these images, the following steps are required [10]:

1. In the first step, a blurring and thresholding filter is applied to the image which makes it possible to differentiate the light spots from the rest of the image.
2. After that, a peak finding algorithm is used to find the centre of each spot individually. Currently, a Gaussian fit algorithm [11] or the method of moments [12] can be used for that.
3. After that, the program identifies the marker in the image, by which the position on the aperture mask can be identified. This is necessary to combine the light spots with the aperture holes in the next step.
4. Finally, a vector from each spot centre through the corresponding hole in the aperture mask is defined (around 700 in total) after which their intersection point can be calculated, which is the actual position of the LED.

The biggest challenge is in fact to determine the centre of the light spots. Since it is not accurate enough to calculate the spot centre to within one pixel, methods must be used that can determine the centre in the subpixel area. With the current state of the art, two different algorithms are used for this. While the Gaussian algorithm is precise and slow, the Moments algorithm is fast and less precise. Since the accuracy in determining the spot's centre position directly affects the finally calculated LED position, we want an algorithm that is as accurate as possible. Nevertheless, it should also be as fast as possible in order not to lose measuring speed. For that reason, a new approach to determine the spot centre with a Convolutional Neural Network is presented in this work.

3.1. Factors influencing the measurement uncertainty

One of the biggest factors is the distance between the LED-Target and the detector. It depends on the ratio of the camera sensor size to the distance. With increasing distance, the changes in the image get smaller, which makes them more difficult to detect. The measurement uncertainty increases because the influence of the disturbance variables increases as a result (*a full explanation with formulas and graphics will follow in the final paper*). Other uncertainty influences such as temperature changes have not yet been investigated in detail. Another major influencing factor is the available number of pixels per spot. The more pixels a spot is imaged with, the more precisely its centre can be determined. Another aspect is the sensor noise. Especially when the intensity of the light spot decreases, the noise of the sensor has a stronger effect on the measured value. In terms of the Neural Network, the number of simulated images, the NN is trained on, affects the measurement uncertainty. The more data available, the more consistent the prediction of the Neural Network is, as experimental trials have shown (*results will be included in the final paper*).

3.2. *Convolutional Neural Network approach*

Since the MAPS images have a high resolution of 3296 x 2472 pixel with a depth of 8-Bit, a lot of hardware resources, especially video RAM is required. Downscaling is not an option, since the relevant information is then lost and the CNN will not learn the relation between input and output, as preliminary experiments have shown. Dividing the image into smaller sections is also not possible, otherwise, the relation between the spots and the aperture holes is lost. Not to mention that the number of combined spots is directly related to the accuracy in LED position determination. For that reason, the task is divided into two parts.

Each MAPS image is an arrangement of light spots, as shown in **Figure 2**. The idea is to first use a CNN to determine the centre of each of the light spots in the image, replacing the currently used algorithms. This allows the entire image to be divided into approx. 700 smaller images, reducing the necessary hardware resources. The CNN is trained on simulated light spots, of which the centre position is known, instead on real spots. Since we don't know the spot centre of the real light spots, only the currently used algorithms can calculate them. When we use this information to label the spots and train the CNN on that data, it will never outperform the traditionally used algorithms. It is important, that the CNN achieves a subpixel accuracy in determining the spot centre, just like the algorithms used so far. But the goal is to perform better than these.

In the second step, another CNN is designed that takes the spot centres as input and returns the LED-Target position. This approach aims to investigate whether systematic errors in the classical algorithms can be avoided and whether an increase in the accuracy of the LED position determination can be achieved.

3.3. *Light spot simulation*

The light spot simulation is a decisive part of this research, as the simulated dataset is what the Neural Network learns from. They should be as close as possible to the real spot images to guarantee a good performance of the CNN later. The simulation is realised in a custom-made python program, where the image and spot size can be set, as well as random noise and different filters. The number of images generated depends on how many spot centre positions are given to the simulation. These positions are interpreted as offset values to the centre of the simulated image.

As a first approach, the light spots are approximated in the simulation by using a two-dimensional Gaussian function according to Equation (1) [13]:

$$f(x,y) = A \exp\left(-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)\right) \quad \textbf{(1)}$$

where x_o and y_o represent the centre coordinates of the spot. By adding an offset value to these, for example, 0.1 it is possible to move the centre in both directions with subpixel accuracy. The smaller the offset steps, the more simulated spots are created and the better the prediction accuracy of the CNN later on. That also allows generating a huge batch of data which is crucial for CNN's training phase.

For the first experiment, three data sets with different noise levels (0 %, 1 % and 4 %) are generated, each with 100k samples. 4 % noise represents the real images ideally, no noise is an optimal image and 1 % is chosen because it is in between. The

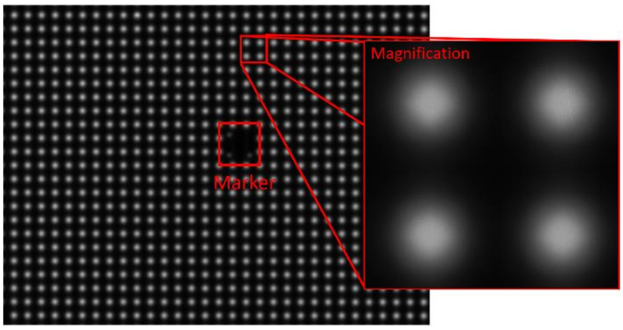


Figure 2: Actual image taken by MAPS at a LED distance of 260 mm from the sensor. The magnification shows four of the light spots, where the gaussian distribution is more clearly visible. The two by two spot sized recess in the middle, with some smaller light spot, is the so-called *marker*. It is used to determine which section of the aperture mask is shown in the image.

difference between the noise levels can be seen in **Figure 3**. To keep this experiment simple, the spot centre is only shifted in positive x-direction by 0 to 1 pixel with a step size of 0.00001 pixels. In addition, random noise of 0 - 10 % of the step size is added to the offset values, which approximate the simulated spots to the real ones. Finally, both the moments and the Gaussian algorithm are applied to each image to calculate the spot centres, which serve as comparative values. The aim of this experiment is only to determine the CNN's subpixel accuracy and compare it to the other algorithms.

Additional experiments and results are presented in the final paper. These will include the simulation of different spot sizes and a greater variety of spot centres, which simulates different distances and angles between the LED-Target and the detector. We will also present the simulation of a complete measured image, determining all the spot centres from which the position of the LED-Target is calculated. Finally, we will use real measured images from which we calculate the position of the Target and compare the accuracy with the state of the art algorithms.

3.4. *CNN preparation and training*

The CNN used in the first experiment is a variation of what Rosenfelder presented in his work [14]. It is an image regression CNN, which means it takes an image as input and outputs a numeric value. In our case, we use a spot image as input and get the spot centre offset in return. The Network is realised in the programming language Python.

The data preparation is kept simple, as complex data augmentation does not bring great advantages, as preliminary research has shown. Only normalisation is applied to the images. The CNN is trained on each of the three data sets individually. The images are split into 70 % training data, 20 % validation data and 10 % test data. An additional data set including images with randomly generated centre offsets are used to compare the

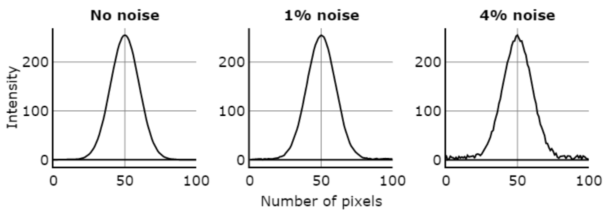


Figure 3: Intensity diagram of the simulated spots with different random noise levels of 0 %, 1 % and 4 %. 4 % noise is close to the real camera sensor noise and therefore ideal to represent the spots. The y-axis represents the pixels (8-Bit) intensity (arbitrary units), the x-axis shows the number of pixels in the spot.

performance of the CNN with the traditional algorithms. The training is set for 200 epochs and a batch size of 128, with early stopping to reduce training time and avoid overfitting.

4. Experimental results

Training the Neural Network in the first experiment on the data sets takes around 50 epochs before the early stopping call back interrupts the learning phase. Applying the trained Network on the test data set with randomly generated offset values gives the results shown in **Table 1**.

At 4 % noise, the CNN outperforms both the Moments and the Gaussian algorithm. With a standard deviation of 0.0411, it is twice as accurate as the Moments algorithm and 1.6 times more accurate as the Gaussian. As the data set with the 4 % noise best represents the real data, this is the most important result. The performance at 1 % noise is similar which means, that the CNN can learn the relevant relations between in- and output without being much affected by the noise in the image. The other two algorithms performed much better at this lower noise level.

At 0 % noise, the CNN performs the worst, with a standard deviation of 0.1603 pixels. This is due to great outliers in some of the prediction values. Since we cannot simply remove these, the performance cannot be improved either. Only if a method is found to avoid the CNN predicting some very inaccurate values, or if a way is found to statistically remove them from the results, is it possible to improve performance.

The Moments algorithm performs much better on this data set than the Neural Network with a standard deviation of 0.0360 pixels. The Gaussian algorithm shows of course no deviation since we are using a Gaussian fit to an ideal Gaussian distributed spot.

Table 1: Comparison of the different methods for determining the spot centre position offset in simulated images. Each method was evaluated by the test data set described in section 3.3. The table shows the standard deviation of the pixel inaccuracy for each method at different noise levels.

	0 % noise	1 % noise	4 % noise
CNN	0.160273	0.041831	**0.041051**
Moments	0.035977	0.029783	0.085795
Gaussian	**0**	**0.016018**	0.067452

5. Conclusion

In this article, a novel approach to determining the LED-Target position from MAPS images using a Convolutional Neural Network was presented. Proprietary simulation software was introduced which was used to create artificial MAPS images. The light spots created by the software have been proven to be very similar to the real ones when a two-dimensional Gaussian function, noise and filters are used (*will be in the final article*). A method was introduced to overcome the challenge of training the Neural Network on high-resolution images by dividing the problem into two tasks. The first CNN was trained on artificial light spots generated by the simulation software, while the second one was trained on the spot's centres and LED-Target position (*final paper*). In the first experiment, we aimed to outperform the traditionally used algorithms with the CNN at different noise levels. This was achieved in the experiment that comes closest to reality at 4 % noise. The CNN predicted the light spot centre offset with a standard deviation of 0.041051 pixels, twice as accurate as the Moments algorithm and 1.6 times more accurate as the Gaussian.

In conclusion, the simulation is capable of representing the real images taken by MAPS. The Neural Network trained on these data outperforms the mathematical methods in the presented experiment. By determining the spot centre more precisely, an increase in MAPS' total accuracy is to be expected. In future work, both the simulation and the CNN design can be improved. Finally, more data sets can be generated using a broader spectrum of simulation parameters like noise to increase the performance of the CNN further.

References

[1] H. Schwenke, W. Knapp, H. Haitjema, A. Weckenmann, R. Schmitt, and F. Delbressine, "Geometric error measurement and compensation of machines—An update," *CIRP Annals*, vol. 57, no. 2, pp. 660–675, 2008, doi: 10.1016/j.cirp.2008.09.008.

[2] P. S. Huang and J. Ni, "On-line error compensation of coordinate measuring machines," *International Journal of Machine Tools and Manufacture*, vol. 35, no. 5, pp. 725–738, 1995, doi: 10.1016/0890-6955(95)93041-4.

[3] H. Kunzmann, T. Pfeifer, R. Schmitt, H. Schwenke, and A. Weckenmann, "Productive Metrology - Adding Value to Manufacture," *CIRP Annals*, vol. 54, no. 2, pp. 155–168, 2005, doi: 10.1016/S0007-8506(07)60024-9.

[4] S. Sartori and G. X. Zhang, "Geometric Error Measurement and Compensation of Machines," *CIRP Annals*, vol. 44, no. 2, pp. 599–609, 1995, doi: 10.1016/S0007-8506(07)60507-1.

[5] U. Bielke *et al.*, "Photogrammetric approach for measuring the position of a Tool-Center-Point using pinhole optics," *(unpublished)*, 2022.

[6] H. Tayara, K. G. Soo, and K. T. Chong, "Vehicle Detection and Counting in High-Resolution Aerial Images Using Convolutional Regression Neural Network," *IEEE Access*, vol. 6, pp. 2220–2230, 2017, doi: 10.1109/ACCESS.2017.2782260.

[7] J. C. Pyo *et al.*, "A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery," *Remote Sensing of Environment*, vol. 233, no. January, p. 111350, 2019, doi: 10.1016/j.rse.2019.111350.

[8] X. Ren *et al.*, "Regression Convolutional Neural Network for Automated Pediatric Bone Age Assessment from Hand Radiograph," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 5, pp. 2030–2038, 2019, doi: 10.1109/JBHI.2018.2876916.

[9] G. Ian, B. Yoshua, and A. Courville, "Deep learning adaptive computation and machine learning," *Prmu*, pp. 1–801, 2016, [Online]. Available: www.deeplearningbook.org

[10] Uwe Bielke *et al.*, "Simulation and accuracy evaluation of a new 3D photogrammetric position measurement system," *(unpublished) Euspen's 22nd International Conference & Exhibition*, vol. 22, Jun. 2022.

[11] N. Hagen and E. L. Dereniak, "Gaussian profile estimation in two dimensions," 2008.

[12] K. O. Bowman and L. R. Shenton, "Estimator: Method of Moments," *Encyclopedia of statistical sciences*. John Wiley & Sons, 1998.

[13] G. L. Squires, *Practical Physics*. Cambridge University Press, 2001. doi: 10.1017/CBO9781139164498.

[14] M. Rosenfelder, "Transfer Learning with EfficientNet for Image Regression in Keras - Using Custom Data in Keras," *rosenfelder.ai*. rosenfelder.ai, 2020.