Real-time Signal identification in Photon Science Imaging

Daniel Becker
University of Applied Science
Berlin, Germany
Email: beckerd@htw-berlin.de

Achim Streit
Steinbuch Center for Computing
Karlsruhe Institute of Technology, Germany
Email: achim.streit@kit.edu

Abstract - The atomic structure of tiny samples like proteins or viruses can be explored using X–ray microscopy. An X–ray laser flash illuminates crystallized samples and the scattered light is captured by a detector device in form of an image. Due to limitations in the experimental setup, the images are contaminated by noise from various sources. In this article, we discuss different sources of noise and propose a series of techniques for identifying useful images in real–time already during the data taking period. The outcome of our algorithm is then compared to predictions obtained by the tools currently in use. We also review our previous work on neural networks for data categorization. In addition, we discuss the Block-Matching 3D (BM3D) algorithm which is considered as state of the art for image de-noising.

Keywords - image processing, photon science, big data, signal processing, X–ray microscopy, nanocrystallography.

I. INTRODUCTION

Nanocrystallography is a subdivision of photon science where the structure of tiny samples is explored by illuminating them with X–rays. The samples are crystallized because then the reflected X–rays show intense peaks, the so–called Bragg spots. From the Bragg spots a three-dimensional image of the internal structure of the samples can be reconstructed.

Fig. 1 shows the setup of the LCLS experiment at Stanford[1]. A jet stream is used to transport the samples across the light source. The light source is a pulsed X–ray laser beam producing 120 flashes per second. For each flash an image is taken. However, since it is not possible to synchronize the stream of samples with the laser pulses only approx. 5% of the samples are fully illuminated in a way that allows further analysis. Consequently, up to 95% of the data are useless for further research [2]. Currently, all data produced by the detector are stored offline.

But this will not be an option in the future. For example, the European X–ray Free Electron Laser (XFEL) will operate at the much higher rate of 27,000 images per second [3]. To achieve the full potential of this experiment, new solutions have to be explored in order to reduce the amount of data actually stored for later analysis.

In this article, we focus on the identification of actual data within an image, namely determining the position of the Bragg spots with a view on further pursuing data reduction in near–real time. Bragg spot identification is made difficult due to noise smeared non–uniformly over the image due to the diffraction of laser light by the buffer solution. In addition, there is a noise contribution because of stuck or broken pixels in the detector device. In this article, we present three connected techniques for localizing Bragg spots. Firstly, noise removal based on a convolution technique. Secondly, edge detection using a technique, well known from image processing and, thirdly, an algorithm that is suggested for finding local clusters within an image. LCLS data from three different samples are used to verify our approach. The results are compared with results obtained by the Cheetah toolkit, the standard software for Bragg spot finding for data taken at LCLS [4].

We review our previous work on neural networks for data categorization. In addition, we discuss the Block-Matching 3D (BM3D) [5] algorithm which is considered as state of the art for image denoising [6]. We are interested in the question whether BM3D can also be used for noise removal in photo science.

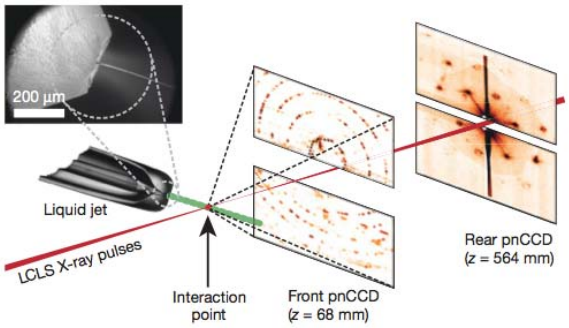


Fig. 1. Nanocrystallography. X-ray pulses from a free-electron laser (FEL) interact with nanocrystals flowing in a buffer solution. The detector records the resulting diffraction patterns [7].

A. Related Work

In the article Crystalline object evaluation by image processing [8] Billingsley et al. use an edge detection algorithm to verify whether the crystallization of protein samples was successful. During the growth process of the crystal, images are taken and an edge detection is performed. This information is used for finding connected lines within an image which, in turn, indicate whether a crystal will generate exploitable diffraction patterns. Within our analysis, we use the same edge detection algorithm.

The currently used solution for pre–selecting images at the LCLS experiment is called "Cheetah" [4]. It uses an algorithm called "hitfinder" to identify Bragg spots in given images. The algorithm looks for a certain amount of connected pixels above a predefined intensity threshold. In addition, several optimization techniques have been incorporated to compensate for background noise. The results of the "Cheetah" software are taken as a reference for rating our algorithm.

In [9] we introduced a neural network for identifying useless data already when they are taken.

Neural networks are often composed of three layers (see Fig. 2). The first layer is the input layer to which the data are fed. This layer consists of one or more neurons. The hidden layer typically consists of several neurons. The weights of the connections between the input layer, the hidden layer and the output layer are trained in such a way that the values of the output neurons reproduce the expected values for a given set of input values quite closely.

An important property of a neural network is generalization, i.e. the ability to correctly categorize new input data.

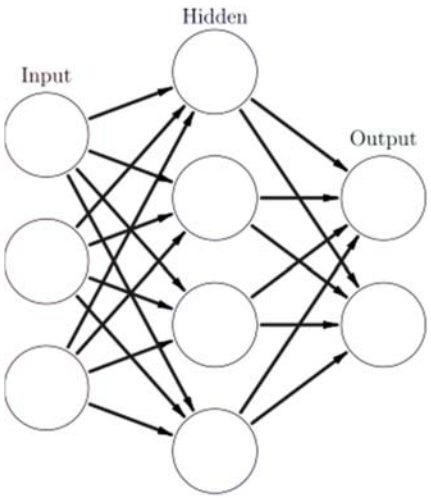


Fig. 2. Schema of a typical neural network.

In our case, a neural network is used to distinguish useless data from images showing Bragg spots. The readout I_i of the i–th pixel is normalized to the unit interval, $0 \le I_i \le 1$. For every diffraction image three quantities are determined,

- the maximum intensity,

$$I_{\max} = \max_{0 \le i \le n} (I_i) \tag{1}$$

- the mean intensity,

$$I_{\mathrm{mean}} = \frac{1}{n} \sum_{i=1}^{n} I_i \tag{2}$$

- and the standard deviation

$$\Delta I = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (I_i - I_{\mathrm{mean}})^2} \tag{3}$$

where n denotes the number of pixels.

These basic data can be extracted in a fast and efficient way. The corresponding neural network consists of three input neurons and two output neurons. No hidden layer is needed (see Fig. 3).

The network is trained by a series of blank and "hit" images. This process adjusts the weights of the connections between the neurons.

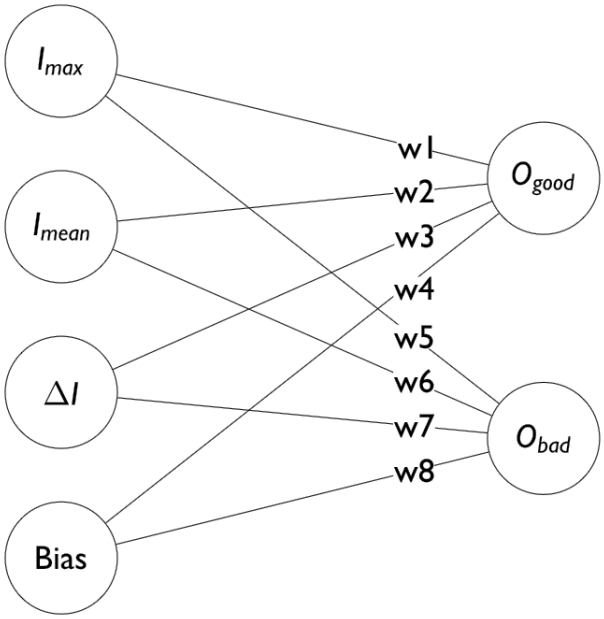


Fig. 3. Schema of our neural network [9]

During an experimental run the characteristic properties of the transportation liquid can change [4]. These changes are easily taken into account by properly adjusting the network weights.

We found out that our network is able to correctly recognize up to 95% of the samples. However, this recognition rate decreases rapidly as the noise within the image increases.

II. DATA & METHODS

A. Experimental Background

The experimental setup can be seen in Fig. 1. The data taken by the experiment go through a chain of processes. These are illustrated in Fig. 4.

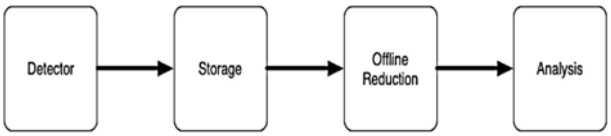


Fig. 4. The flow of data taken at LCLS..

The experiment at LCLS runs at an image repetition rate of 120Hz. This means that the laser produces 120 flashes per second and for each flash an image is taken by the detector. The detector produces gray-scale images with a resolution of 2.3 Megapixel leading to an image size of about 4 Megabytes. Consequently, ~1.8 TB of data are taken per hour.

Currently all data are stored offline prior to any analysis. After storing the data, an offline reduction is done using the Cheetah toolkit (see Sec. I-A1). The resulting subset is then used for the actual analysis.

As mentioned in Sec. I, this will not be an option for new generations of experiments. The European XFEL experiment, for example, will operate at an image repetition rate of 27,000 Hz. Currently, the resolution of the detector device to be used at XFEL is not known. Nonetheless, if we assume that it has at least the same resolution as the detector used at LCLS, this will increase the amount of data taken by factor 225, leading to a volume of about 405 TB/h. The consequence of this is that storing all data offline is no longer an option.

B. Test Data

Diffraction data from three crystallized samples are analyzed:

- the protein Cathepsin B (CatB) [10],
- the 5-Hydroxytryptamine receptor 2B (5HT-2B) [11],
- the granulovirus polyhedron (GV) [12].

These samples are giant molecules. As a consequence they are hard to crystallize for X–ray microscopy [2]. This is especially true for the receptor and the virus samples. Because of that, additional noise can be introduced due to the complex structure of the crystals.

For each sample, 25 indexable and 25 non–indexable images are selected. An image is indexable [13] if, roughly speaking, a Fourier series index can be assigned to each Bragg spot of a diffraction pattern. Bragg spots and the electron density of a sample are related (in lowest order perturbation theory) by a Fourier transformation, see e.g. [14].

C. Geometry

The detector device is composed of multiple panels (see Fig. 5). The panels are organized into four quadrants, which are able to slide in and out relative to the center of the detector.

However, this results in varying coordinates for each detector pixel within an image. The geometric details are stored in a separate file for each experiment.

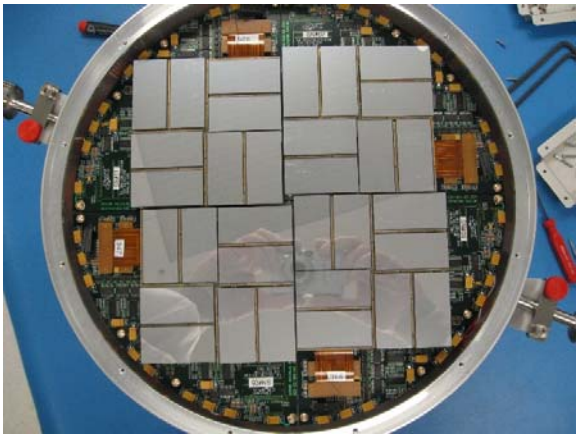


Fig. 5. The CSPad detector device used at the LCLS experiment at Stanford.

D. Data Normalization

The readout of the pixels is corrected for nonphysical data: The detector collects 14-bit unsigned integer data per pixel [15]. Negative or very large (>16,000) readout numbers indicate broken or stuck pixels and are therefore set to zero.

Let $I_i^{(raw)}$ be the readout number of the i–th pixel of an analyzed image. The quantity

$$I_i = I_i^{(raw)}\theta(I_i^{(raw)})\theta(16{,}000 - I_i^{(raw)})$$

represents the intensity of the i–th pixel. θ(p) is the unit step function (zero for negative p, one for non–negative p). The intensity is normalized to the interval, $0 \leq I_i \leq 16{,}000$.

Each detector pixel i can also be identified by two integers (x,y) characterizing the position of a pixel in the detector plane where $0 \leq x,y < \dim(I)$.

E. Average Subtraction

Laser light is always diffracted when propagating through the buffer solution resulting in an overall background. This background can be removed to a large extend from an image by analyzing a series of known blank images and determining an average noise level for each pixel of the image.

To estimate the average background noise, N=500 images without diffraction information (blanks) are analyzed in advance and for every pixel i the mean noise

$$I_i^{(\mathrm{noise})} = \frac{1}{N} \sum_{n=1}^{N} I_{n,i}^{(\mathrm{blank})} \tag{4}$$

is determined, where $I_{n,i}^{(\mathrm{blank})}$ denotes the intensity of the i–th pixel in the n–th blank image (corrected for nonphysical readout data, see Sec. II-D.

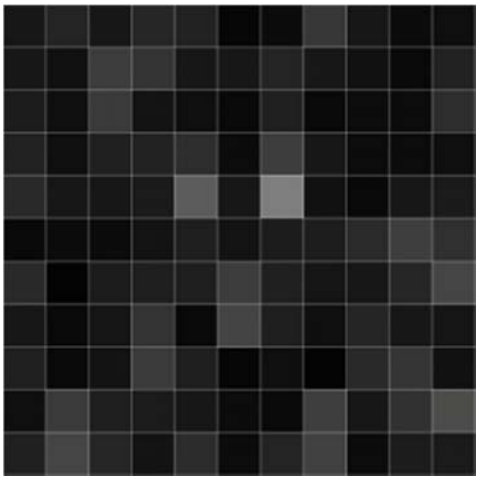
The influence of background effects is explored by considering the corrected intensity

$$\widetilde{I}_i = (I_i - I_i^{(\mathrm{noise})})\theta(I_i - I_i^{(\mathrm{noise})}) \tag{5}$$

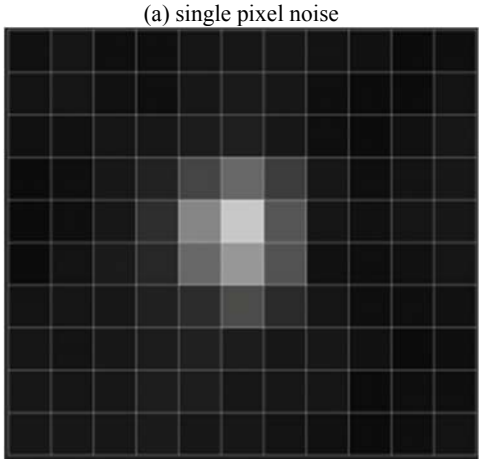
It should be noted that the noise subtraction depends on the kind of fluid used to transport the crystals. In this article, it is assumed that the background is static during a measurement period.

F. Noise Reduction

Diffraction images contain noise from different sources. In this section we concentrate on sharply localized noise affecting only single detector pixels. In contrast, Bragg spots are usually distributed over several pixels. An example for noise as well as a valid Bragg spot can be seen in Fig.6.



(a) single pixel noise



(b) Bragg spot

Fig. 6. Example of single pixel noise (6a) and a Bragg spot (6b).

The probability for misidentifying a single spot pixel as a Bragg spot can be reduced considerably by distributing its intensity over the neighboring pixels by using convolution techniques well known in image processing [16]. Let us introduce the three–dimensional matrix

$$K = \frac{1}{9}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

By convoluting an image I with K

$$I^{(\mathrm{noisexreduced})} = I * K$$

where the convolution (*) is defined by

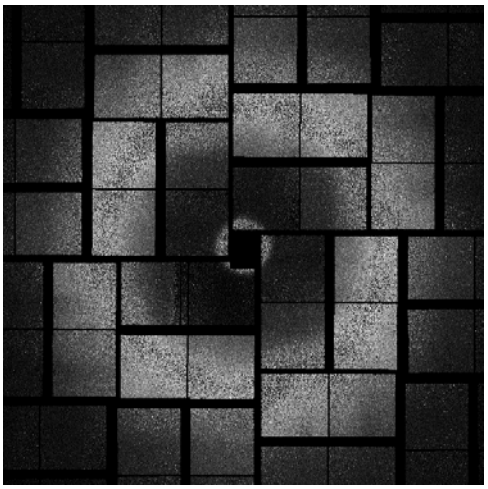
$$I_{x,y}^{(\mathrm{noisexreduced})} = \sum_{m=1}^{\dim K} \sum_{n=1}^{\dim(K)} I_{x-m+2,y-n+2} K_{m,n},$$

the strength of a single pixel noise at the position (x,y) can be reduced.

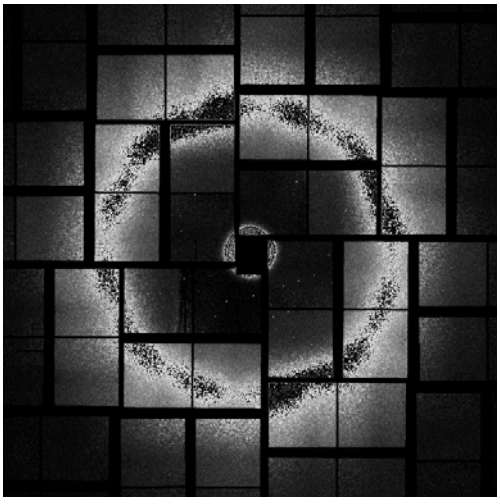
For consistency, the matrix I is set to zero beyond its boundaries, i.e. $I_{x,y} = 0$ for $x,y < 0$ and $x,y \geq \text{size}(I)$.

Note that there is no overall intensity loss due to the convolution operation as the normalization of the matrix K is chosen such that the sum of its matrix components is one. After convolution, the intensity of isolated single–pixel spots is damped considerably and approximately comparable with the average noise in the neighborhood.

An example of the influence of convolution is shown in Fig. 7. As can be seen, single pixel noise is drastically reduced and consequently, spots composed of more than one pixel are now standing out against the background. This is visible in the area around the center of the image.



(a) before



(b) after

Fig. 7. Example of an image before (7a) and after (7b) convolution with the matrix K.

G. Edge Detection

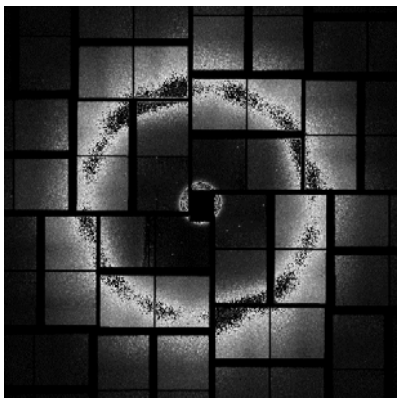
Bragg spots differ in size, shape, and intensity. However, what many spots seem to have in common is the shape of their intensity distribution. This characteristic feature can be harnessed for identifying Bragg spots by analyzing their boundaries. A standard technique for detecting edges within an image is using the horizontal and vertical Sobel operators [17].

$$S_h = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, S_v = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}$$

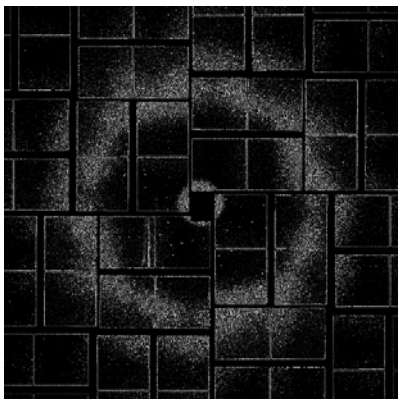
They are convoluted with an image to determine horizontal and vertical intensity changes. The quantity

$$\Delta I_{x,y} = \sqrt{\left(I^{(\text{noisereduced})} * S_h\right)^2_{x,y} + \left(I^{(\text{noisereduced})} * S_v\right)^2_{x,y}}$$

can be considered as a measure for the strength of the local intensity change at the pixel position (x,y) of the image I. Large values for $\Delta I_{x,y}$ indicate an edge at (x,y).



(a) before



(b) after

Fig. 8. Example of an image before (8a) and after (8b) application of edge detection.

Fig. 8 shows the results of the application of edge detection to a noise–reduced image. It can be seen that the spots around the center now stand out even more clearly. Furthermore, the previously still noisy water halo has been damped as well, making the spots loom clearer in comparison.

```
1:  procedure CLUSTERDETECTION(image)
2:    clusterMatrix ← dim(image)
3:    maximumDistance ← 10
4:    brightnessThreshold ← 400
5:    for x = 0; x < size(image); x++ do
6:      for y = 0; y < size(image); y++ do
7:       for i = x - maximumDistance; i + maximumDistance; i++ do
8:         for j = x - maximumDistance; j + maximumDistance; j++ do
9:          if image[x][y] >brightnessThreshold
               && distance(image[x][y],clusterMatrix[x][y]) then
10:          clusterMatrix[x][y] ← clusterMatrix[x][y] + 1
11:         end if
12:        end for
13:       end for
14:     end for
15:   end for
16:  end procedure
```

Fig. 9. Cluster detection in pseudo–code

H. Cluster Detection

After applying an edge detection, Bragg peaks can be viewed as a cluster of bright, connected pixels. This means a simple cluster finding algorithm can be used to find connected pixels within the image, see Fig. 9.

First, an empty matrix is created using the dimensions of the analyzed image (cluster matrix). Then, each pixel of the original image is examined. For each of these pixels, the algorithm looks at each surrounding pixel within a defined distance. If the intensity of one of these adjacent pixels is higher than a defined threshold, the intensity of the corresponding original pixel in the cluster matrix is increased.

The cluster detection algorithm is not designed for rating the size and shape of the Bragg spots found. Since the lack of insights from different areas, e.g. from theoretical models, it is not possible to make a clear decision, whether a candidate is too small or large to represent an actual Bragg spot. Therefore, only clusters which are obviously too big to represent a Bragg spot are removed. This threshold has been set to 1000 pixels.

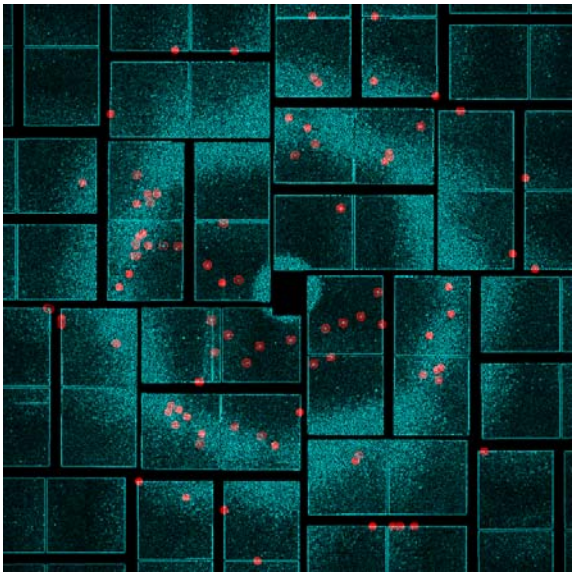
Once there is a better awareness about the actual boundaries for spot sizes, the cluster algorithm can be extended so that spot candidates violating the limits are removed.

The geometry of the individual panels within the image is known. Therefore it is possible to remove falsely found spots at the edges of a panel.

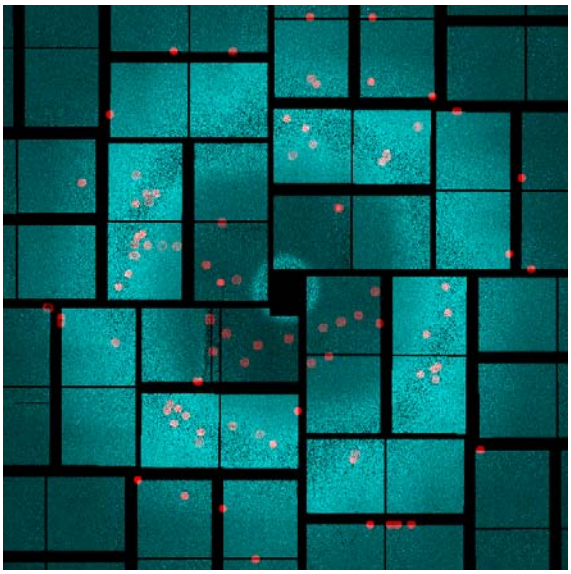
The image from Sec. mbox II-H mbox containing the clusters found by our algorithm has been updated by removing all spots on panel edges. The resulting image can be seen in Fig. 11.

I. Optimization

1) Noise removal and edge detection can be applied almost in parallel, with a short delay in between. Let us assume, that noise reduction has already been done to a 3x3 block at the top left corner of the image. The top left pixel can be convoluted with the image kernels used for edge detection, since these kernels also have a size of 3x3.



(a) Image with applied noise reduction & edge detection



(b) Image without optimization

Fig. 10. Clusters found overlaid on the optimized image (10a) and default image (10b).

This is possible, because all of the adjacent pixels necessary for the calculation have already been processed by noise reduction. By continuing the noise removal row– and column–wise towards the bottom right corner of the image, an edge detection can be done in parallel, provided the noise reduction is always 3 pixels ahead of the edge detection in order to ensure that only noise reduced data are used for edge detection.

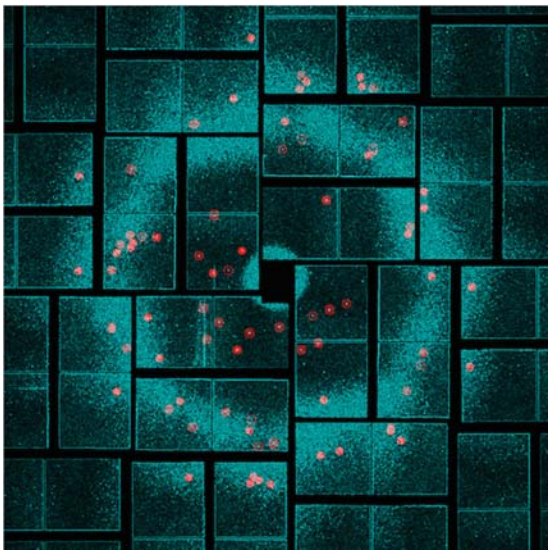


Fig. 11. Clusters found with spots on edges removed.

2) Our algorithms are able to run in parallel. Due to the high amount of data that need to be processed, it is very important to be able to not only process individual images in parallel, but rather divide each image into sub–images that can be processed in parallel, too. This can be achieved by the algorithm proposed in this article as well as by our algorithm based on neural networks as a categorization engine (see Sec.I-A2. Both do not depend on any aspects of the whole image including the details of the geometry. An image is composed of the sub–images of 64 panels. These can be processed independently and once one sub–image has undergone noise removal, the edge detection and the cluster finding can be applied independently of the other images. To decide whether the image can be used for subsequent analysis, it is only necessary to calculate the sum of the spots found in each sub–image.

As an alternative to the Sobel operator, the Scharr operator ([18]) could also be used for edge detection. It defines slightly different kernels for edge detection x and y direction:

$$Scharr_h = \begin{pmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{pmatrix}, Scharr_v = \begin{pmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{pmatrix}.$$

These operators were proposed by Scharr to optimize the rotational symmetry of the operators. They represent an alternative to the Sobel operator currently used for edge detection.

III. CONSIDERATIONS ON BLOCK-MATCHING 3D

One example of a modern noise removal algorithm is the Block-Matching 3D (BM3D) algorithm. The work flow of BM3D is shown in Fig. 13.

The algorithm removes noise in two steps. Fig. 12 shows a typical example where the first step already removes most of the noise and the second step corrects for remaining artifacts, e.g. in the sky.

In the first step, similar sub–images (blocks) are identified within the image. This is achieved by calculating the distance to a reference sub–image. These blocks are then transformed into the Fourier space. To each transformed blocks, a hard thresholding is applied in order to separate noise from signal.

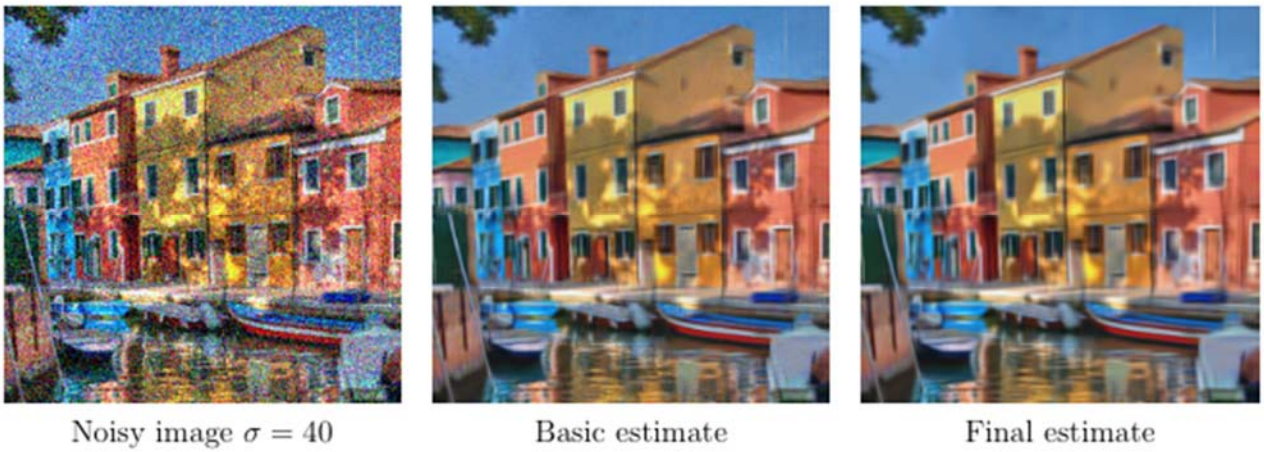


Fig. 12. Example of the BM-3D algorithm at work. From left to right: noisy source image, image after step 1 and image after step 2[19].

Then, the inverse Fourier transformation is applied and estimates about the blocks are gathered. In case of overlapping blocks, the aggregation is not trivial and the calculated hard thresholds are used to determine the more influential block.

In the second step, the results are refined. Here, the blocks that have been processed by step 1 are used again as well as a version of the blocks before step 1 is applied. This pair of blocks is then transformed into the Fourier space and used as the input for a Wiener filter. The already processed block from step 1 is taken as a representation of the desired signal and the unprocessed block is used as the noisy input. The Wiener filter is used in signal processing to suppress noise within a signal. Then, the Fourier transformation is inverted, estimates about the noise are gathered and then aggregated. This leads to the noise reduced version (final estimate) of the original image.

Both steps require a separation of signal and noise in order to work properly. Bragg spots do not differ strongly in shape from the random noise throughout the image. Consequently, it is not possible separate signal and noise in the Fourier space, since the support of the Fourier transform of narrow signals is wide and strongly overlapping with the Fourier transform of the noise.

For the success of BM3D, it is important that the sub-images can be represented in a way, that the information are separable into noise and signal. If this condition is violated, artifacts are introduced.

Traditional algorithms for removing spot–like noise will not work in Photon science, since they would remove single pixel noise as well as Bragg spots. As can be seen in Fig. 6, a Bragg spot is only a few pixels in size and hardly distinguishable from the single pixel noise.

IV. RESULTS

The proposed techniques in this paper have been applied to three samples (see Sec. II-B). The Bragg spots identified by using our "cluster finder" algorithm have been compared to the results obtained by the "hitfinder" algorithm from the Cheetah toolkit.

Our analysis is split into four classes. One distinctive property is the removal of background photon noise before the images are analyzed. Furthermore, the analysis is applied to indexable and non-indexable images. Non–indexable means that an image is not suited for further research.

The results are shown in Tab. I and Tab. II where no background subtraction is applied and in Tab. III and Tab. IV where background subtraction has been applied. It can be seen that the spots found by hitfinder as well as clusterfinder increases in cases, where background subtraction has been applied. This is due to the additional reduction of noise, which reduces the number of false positives by our algorithm.

V. CONCLUSION

In this article, we introduced techniques which, when combined, enable a detection of the majority of Bragg spots of a diffraction image. To this end, we connected the ideas of noise reduction, edge detection and cluster finding. We compared the spots found by our algorithm with the currently used Cheetah software. Depending on the noise level of the images, we are able to find up to 90% of the spots found by the hitfinder algorithm, as well as additional ones.

The proposed algorithm in this article could also be combined with our approach on using neural networks as a categorization engine (see Sec. I-A2). The proposed network could likewise be extended by a third output neuron. The network could then be trained to classify the images into blanks, most likely hits and "in-between". Images falling into the latter category could then be further investigated upon using the clusterfinder algorithm. This would greatly reduce the amount of data requiring extensive processing and thus save computational resources.

It would also be possible to just apply the noise removal on the images in question and then re–submit those into the network. This way, the number of images a cluster finding has to be done to, would be reduced even further.

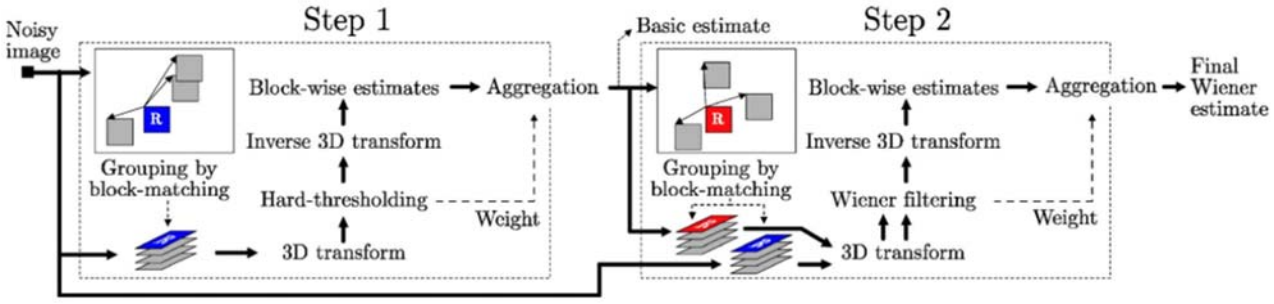


Fig. 13. The work-flow of the Block-Matching 3D algorithm[20].

The main problem in detecting correct Bragg spots is the amount of noise introduced by various sources throughout the experiment. In this article, we explored a convolution technique as well as average subtraction which, together, are able to remove a significant part of the noise. However, it should be further investigated whether there are alternative or additional approaches for reducing the noise even more and, finally, to improve the correct recognition of Bragg spots.

The convolution based techniques for Bragg spot finding presented in this article can be parallelized, in principle, by splitting an image into several parts where a natural split is given by the sub-images collected by the panels of the detector device.

To remove the background noise even further, it should be investigated, if it is feasible to calculate a dynamic brightness threshold for finding bright spots within an image. This could be done by evaluating a time–series of images and adjusting the threshold dynamically.

Whether our cluster finding algorithm might be useful in identifying Bragg spots in near–real time immediately after the data taking period and, moreover, to contribute to damming the flood of data at XFEL, remains to be explored.

ACKNOWLEDGMENT

This work is supported by the Portfolio Extension of the German Helmholtz Association ”Large Scale Data Management and Analysis” [21].

REFERENCES

[1] Y. Ding, A. Brachmann, F.-J. Decker, D. Dowell, P. Emma, J. Frisch, S. Gilevich, G. Hays, P. Hering, Z. Huang et al., “Measurements and simulations of ultralow emittance and ultrashort electron beams in the linac coherent light source,” Physical review letters, vol. 102, no. 25, p. 254801, 2009.

[2] S. Boutet, L. Lomb, G. J. Williams, T. R. Barends, A. Aquila, R. B. Doak, U. Weierstall, D. P. DePonte, J. Steinbrener, R. L. Shoeman et al., “High-resolution protein structure determination by serial femtosecond crystallography,” Science, vol. 337, no. 6092, pp. 362–364, 2012.

[3] R. Klanner, J. Becker, E. Fretwurst, I. Pintilie, T. Pöhlsen, J. Schwandt, and J. Zhang, “Challenges for silicon pixel sensors at the european xfel,” Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 730, pp. 2–7, 2013.

[4] A. Barty, R. A. Kirian, F. R. Maia, M. Hantke, C. H. Yoon, T. A. White, and H. Chapman, “Cheetah: software for high-throughput reduction and analysis of serial femtosecond x-ray diffraction data,” Journal of Applied Crystallography, vol. 47, no. 3, pp. 1118–1131, 2014.

[5] J. Ellenberge and C. E. Chow, “Noise reduction in digital imaging-an exploration of the state of the art,” University of Colorado at Colorado Springs, Multimedia Computing and Communications, 2010.

[6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Bm3d image denoising with shape-adaptive principal component analysis,” in SPARS’09-Signal Processing with Adaptive Sparse Structured Representations, 2009.

[7] H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall et al., “Femtosecond x-ray protein nanocrystallography,” Nature, vol. 470, no. 7332, pp. 73–77, 2011.

[8] J. Billingsley, K. Kawabata, M. Takahashi, K. Saitoh, M. Sugahara, H. Asama, T. Mishima, and M. Miyano, “Crystalline object evaluation by image processing,” Sensor Review, vol. 28, no. 2, pp. 143–149, 2008.

[9] D. Becker and A. Streit, “A neural network based pre-selection of big data in photon science,” in Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth International Conference on. IEEE, 2014, pp. 71–76.

[10] L. Redecke, K. Nass, D. P. DePonte, T. A. White, D. Rehders, A. Barty, F. Stellato, M. Liang, T. R. Barends, S. Boutet et al., “Natively inhibited trypanosoma brucei cathepsin b structure determined by using an x-ray laser,” Science, vol. 339, no. 6116, pp. 227–230, 2013.

[11] J. M. Perkel, “Decoding protein structure, one femtosecond at a time,” 2014.

[12] E. Chiu, F. Coulibaly, and P. Metcalf, “Insect virus polyhedra, infectious protein crystals that contain virus particles,” Current opinion in structural biology, vol. 22, no. 2, pp. 234–240, 2012.

[13] F. R. Maia, C. Yang, and S. Marchesini, “Compressive auto-indexing in femtosecond nanocrystallography,” Ultramicroscopy, vol. 111, no. 7, pp. 807–811, 2011.

[14] J. Als-Nielsen and D. McMorrow, Elements of modern X-ray physics. John Wiley & Sons, 2011.

[15] S. Herrmann, S. Boutet, B. Duda, D. Fritz, G. Haller, P. Hart, R. Herbst, C. Kenney, H. Lemke, M. Messerschmidt et al., “Cspad-140k: A versatile detector for lcls experiments,” Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 718, pp. 550–553, 2013.

[16] R. Keys, “Cubic convolution interpolation for digital image processing,” Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 29, no. 6, pp. 1153–1160, 1981.

[17] I. Sobel and G. Feldman, “A 3x3 isotropic gradient operator for image processing,” a talk at the Stanford Artificial Project in, pp. 271–272, 1968.

[18] H. Scharr, “Optimal operators in digital image processing,” Ph.D. dissertation, 2000.

[19] M. Lebrun, “An analysis and implementation of the bm3d image denoising method. image processing on line (2012).”

[20] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-d transform-domain collaborative filtering,” Image Processing, IEEE Transactions on, vol. 16, no. 8, pp. 2080–2095, 2007.

[21] J. van Wezel, A. Streit, C. Jung, R. Stotzka, S. Halstenberg, F. Rigoll, A. Garcia, A. Heiss, K. Schwarz, M. Gasthuber et al., “Data life cycle labs, a new concept to support data-intensive science,” arXiv preprint arXiv:1212.5596, 2012.

TABLES I
FOUND SPOTS IN INDEXABLE IMAGES

	Total spots: hit finder	**Total spots: cluster finder**	**Spots in common [1]**	**Spots: hit finder only [1]**	**Spots: cluster finder only [2]**
CatB	937	1.532	89%	11%	48%
SHT-2B	1,400	2.169	68%	30%	59%
GV	1,180	1.603	42%	57%	60%

[1]Compared to hitfinder.
[2]Compared to cluster finder.

TABLE II.
FOUND SPOTS IN NON-INDEXABLE IMAGES

	Total spots: hit finder	**Total spots: cluster finder**	**Spots in common [1]**	**Spots: hit finder only [1]**	**Spots: cluster finder only [2]**
CatB	0	66	0%	0%	100%
SHT-2B	827	1,768	68%	30%	59%
GV	2,015	1,287	45%	54%	53%

[1]Compared to hitfinder.
[2]Compared to cluster finder.

TABLE III.
FOUND SPOTS IN INDEXABLE IMAGES WITH AVERAGE SUBTRACTION

	Total spots: hit finder	**Total spots: cluster finder**	**Spots in common [1]**	**Spots: hit finder only [1]**	**Spots: cluster finder only [2]**
CatB	937	1427	90%	10%	42%
SHT-2B	1,400	2.169	72%	27%	58%
GV	1,180	1,093	43%	56%	19%

[1]Compared to hitfinder.
[2]Compared to cluster finder..

TABLE IV.
FOUND SPOTS IN NON-INDEXABLE IMAGES WITH AVERAGE SUBTRACTION

	Total spots: hit finder	**Total spots: cluster finder**	**Spots in common [1]**	**Spots: hit finder only [1]**	**Spots: cluster finder only [2]**
CatB	0	22	0%	0%	100%
SHT-2B	827	1,768	59%	39%	68%
GV	2,015	1,204	33%	67%	14%

[1]Compared to hitfinder.
[2]Compared to cluster finder.