Robust impurity detection and tracking for tokamaks

C. Cowley ^(b), ^{*} P. Fuller, Y. Andrew ^(b), L. James ^(b), and L. Simons ^(b) Blackett Laboratory, Imperial College, London SW7 2AZ, United Kingdom

M. Sertoli[®], S. Silburn, A. Widdowson[®], and JET contributors[†] *Culham Centre for Fusion Energy, Culham Science Centre, Abingdon, Oxfordshire OX14 3DB, United Kingdom*

> I. Bykov and D. Rudakov General Atomics, San Diego, California 92186, USA

T. Morgan, S. Brons, J. Scholten, and J. Vernimmen DIFFER—Dutch Institute for Fundamental Energy Research, 5612 AJ Eindhoven, The Netherlands

P. Bryant and B. Harris

Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, United Kingdom (JET, DIII-D, and Magnum-PSI Collaborations)

(Received 24 March 2020; accepted 13 July 2020; published 13 October 2020)

A robust impurity detection and tracking code, able to generate large sets of dust tracks from tokamak camera footage, is presented. This machine learning–based code is tested with cameras from the Joint European Torus, Doublet-III-D, and Magnum-PSI and is able to generate dust tracks with a 65–100% classification accuracy. Moreover, the number dust particles detected from a single camera shot can be up to the order of 1000. Several areas of improvement for the code are highlighted, such as generating more significant training data sets and accounting for selection biases. Although the code is tested with dust in single two-dimensional camera views, it could easily be applied to multiple-camera stereoscopic reconstruction or nondust impurities.

DOI: 10.1103/PhysRevE.102.043311

I. INTRODUCTION

Magnetic confinement fusion in deuterium-tritium plasmas has the potential to be a clean, sustainable, and efficient alternative to current power generation methods. Across the globe numerous tokamaks have been built to achieve sustained nuclear fusion, with the Joint European Torus (JET) [2] and Doublet-III-D (DIII-D) [3] being two of the machines. To achieve ignition in a fusion plasma, high densities, temperatures, and confinement times are required [4]. The central challenge of fusion energy is in achieving these conditions and attempts to do so have been through a multitude of physics and engineering optimizations.

In every operational tokamak exists impurities, which can either exist intrinsically or are extrinsically added for experimental purposes. One class of intrinsic impurity is dust, which is typically a nm-mm particulate created from the tokamak plasma facing components (PFCs). These PFCs include the tokamak walls and divertor, which in current experimental devices are typically composed of carbon, in the form of graphite or carbon-fiber composites, or metals such as tungsten, beryllium, and molybdenum [5,6]. During tokamak operation, high-energy events such as edge localized modes (ELMs), disruptions, and vertical displacement events can load these PFCs with powers up to the order GWm⁻². These events can potentially liberate intrinsic dust and even create more dust via processes such as spallation, vaporization, and erosion of the PFC melt layer [7–9]. In addition to these plasma interactions, mechanical machining of PFCs can also contribute to the creation of dust.

Dust in tokamaks can pose both operational and safety concerns. In particular, dust particles can retain tritium fuel leading to an increase in the in-vessel tritium inventory over time. This is of concern both for safety and for the tritium fuel cycle, and as a consequence future machines have limited the tritium inventory to 700 g [10]. In addition, when dust particles composed of high-mass elements such as tungsten enter the plasma, this can cause sudden cooling of the plasma by increased radiation and a drop in the plasma fusion performance or in extreme cases lead to plasma disruptions [11,12].

In an attempt to understand and mitigate the risks posed by tokamak impurities, the physics of dust in plasmas has become a highly researched area in recent years. At the forefront of this research are dust in plasma simulation codes such as MIgration of GRAINs in fusion devices (MIGRANe) [13], Dust in TOKamaks (DTOKS) [14], DUST-TRACKing (DUSTTRACK) [15], and DUST Transport (DUSTT) [16]. These codes solve differential equations governing the

^{*}cyd.cowley15@ukaea.uk

[†]See the author list of Ref. [1].

evolution of temperature, charge, and velocity of a dust grain in a plasma background [17,18]. One key use for these codes is in determining which physical forces and phenomena play important roles in dust transport in tokamaks. In theory, one of these codes could be run with different forces active, to see which phenomena best replicate experiment. This, however, requires large statistically significant dust track datasets to compare against the models. Such datasets are currently sparse and manually selected. Here the Robust Impurity Detector and Tracker (RIDAT) code is presented, which allows for automatic generation of such datasets from tokamak camera footage.

II. METHOD

A. RIDAT code

The code for RIDAT was developed in *Python*, with aims of detecting significant sets of dust tracks automatically. RI-DAT was also designed to be robust, such that it can be applied to various cameras and impurities. The code can be separated into image processing and dust tracking modules. The image processing module receives a list of image frames from tokamak video footage as an argument and outputs a list of dust grains detected in each frame—along with the dust grain properties. The dust tracking module uses these isolated detected grains as arguments and outputs a list of connected dust tracks spanning multiple frames [19].

The image processing module detects and characterizes dust grains, which can be seen as small bright ellipses or streaks in camera footage. To achieve this, the code works frame by frame, determining a background average for each frame. This is done by averaging the pixel intensities in temporally adjacent frames. This background average is then subtracted from the intensity profile of a frame, and all pixels over a user selected brightness threshold T_B are detected. These bright spots are collected into dust grains by grouping spatially adjacent bright pixels. Two successfully detected grains can be seen in Fig. 1.

Finally, basic properties of the grains are determined, such as position, brightness, length, and width. Position and brightness are found by averaging the *x*-*y* pixel coordinates and normalized intensities of a grain, respectively. The length of a grain is found by taking the longest distance (px) between any two pixels on the grain. The width is then simply the grain area (px^2) divided by length. It is important to note that—depending on camera frame rate—the length of a dust grain can either be indicative of its physical dimensions, or how far it has traveled over the camera exposure period. Because of this, a user can manually select whether RIDAT is in streaking or nonstreaking mode. In the former, RIDAT treats the length as the distance traveled and stores two positions of a grain: one at the start and one at the end of its motion.

The dust tracking module takes the unordered dust grain lists of each frame outputted by the image processing module as an argument. It returns an ordered list of fully connected dust tracks spanning multiple frames. In other words, the purpose of this module is to connect the temporally isolated dust grains found by the image processing module. The core of this module is an algorithm that works frame by frame analyzing



(b)

FIG. 1. (a) An unedited image taken from the DiMES camera in DIII-D. (b) The same still, but with RIDAT's background subtraction applied with a T_B of 35. Detected dust grains are indicated by red circles.

all possible tracks for a dust grain in frame N across frames N + 1 and N + 2. Then the algorithm determines which of these tracks is the most likely real track using a naive Bayesian machine learning classifier. Finally, it repeats the process on the next set of three frames (N + 1, N + 2, and N + 3), joining the new data to any previously detected tracks.

A naive Bayesian classifier is a supervised machine learning classifier that determines the probability of each classifier C_i fitting to data with features f, $P(C_i|f)$, using Bayes's theorem [20]. In the case of RIDAT, the classifier assumes a Gaussian probability density function, and classifies a dust track as either True (corresponding to a likely real track) or False (corresponding to an erroneous joining of unrelated dust grains). The features then consist of the following properties of a potential dust track:

- (i) Mean change in position, $\mu_{\delta R}$ (px).
- (ii) Standard deviation of change in position, $\sigma_{\delta R}$ (px).
- (iii) Mean change in brightness, $\mu_{\delta B}$ (normalized).
- (iv) Mean change in width, $\mu_{\delta w}$ (px).
- (v) Mean velocity angle, μ_{θ} (°).
- (vi) Mean change in velocity angle, $\mu_{\delta\theta}$ (°).

This type of classifier was selected due to its simplicity and ability to determine not only if a track was correct (True), but also the likelihood of correctness. Thus, for a given dust grain in frame N, if there are candidates for the corresponding grain in frame N+1 and N+2, then this classifier can determine the

TABLE I. The features of one correct and one incorrect dust track identified by a user using RIDAT for a shot in Magnum-PSI.

$\mu_{\delta R}$	$\sigma_{\delta R}$	$\mu_{\delta B}$	$\mu_{\delta w}$	$\mu_ heta$	$\mu_{\delta heta}$	С
9.5	2.5	0.4	0.4	$0.0 \\ -52.9$	0.0	True
26.6	20.5	0.7	0.0		48.5	False

most probable correct candidates with a Bayesian probability over a user-defined threshold T_P . The listed features have been chosen as they are the physical properties intuitively thought to discriminate dust tracks from other image features. If a dust grain is traveling at a constant speed, for example, then the mean change in position should be relatively constant. It is important to note, however, that this example of a constant speed is not assumed by the user. In fact, even if all grains were traveling with a constant three-dimensional (3D) speed, the projection onto a 2D camera would not be constant. The only assumption made is that there is some distribution of particle speed; the machine learning algorithm then determines the characteristics of this distribution.

Creation of the training data sets was achieved by a training function that displays three consecutive video frames to the user. The user can select the correct path in frames 2 and 3 for a given grain in frame 1. The dust track features are saved, with a classification of True. Because a given grain only takes one True path, all other potential tracks stemming from the given frame 1 grain are classified False, which allowed the creation of large training data sets quickly. Although this method creates a significant disparity between classifications, it is a disparity mirrored by reality, as the number of possible paths for a grain greatly outweighs the one True path. An example of one True and one False dust track from a training data set from the linear Magnum-PSI machine is shown in Table I.

To test the efficacy of the code, RIDAT was applied to footage from three plasma machines: JET, DIII-D, and Magnum-PSI (M-PSI). The general process of application was first to determine parameters to be used in the code, then to manually create a training data set for each machine, and, finally, to run the tracking code on one shot at a time. Due to the high variation in dust abundance, computational time for tracking varied across machines. The shortest computation time on a four-core computer was less than 1 min for a 10-frame shot in JET, with \sim 1 grain per frame. The most computationally expensive run was on a single 1000-frame batch in M-PSI; taking more than 20 h for \sim 20 grains per frame.

B. Diagnostic setup

Footage from the JET tokamak was captured with a nearinfrared (NIR) Hitachi KP-M1AP protection camera, with a frame rate of 50 Hz. This camera was part of a safety monitoring system, situated on the top half of JET's inner wall, with an angled view toward the divertor [21]. The footage was from a 2018 campaign, during which W dust was occasionally mobilized by the restart of the plasma. Typical footage from the NIR camera is shown in Fig. 2.





FIG. 2. (a) A 3D reconstruction of JET's NIR camera view using the tokamak viewing software Calcam [22]. (b) A still image taken from the same NIR camera.

In addition to JET, footage was collected from the tokamak DIII-D [3]. The camera used for this experiment was a CIDTEK 3710D camera with a frame rate of 5 kHz and a spectral width dependent on the filter applied [23]. The camera was part of a Divertor Material Evaluation Station (DiMES), which has a vertical view of the divertor [24]. This top-down divertor view can be seen in Fig. 3. Shots from DIII-D were captured during an experimental campaign studying





(b)

FIG. 3. (a) A 3D Calcam reconstruction of the DiMES view in DIII-D [22]. The gray tiles indicate the divertor viewed from above. (b) A still image taken from the DiMES visible camera.

turbulent transport and ELM control [25]. Dust grains observed in these videos were most likely composed of intrinsic graphite, as this is the main material used for PFCs in DIII-D.

The final machine used for experimental data was M-PSI, a linear plasma generator located in the Dutch Institute for Fundamental Energy Research. Camera footage collected was from a Phantom V12.1 visible fast imaging camera located in one of the side camera ports of the cylindrical housing, shown in Fig. 4. Footage was collected from over a dozen shots with a frame rate spanning from 400 Hz to 1 kHz [26]. The footage for this research was captured during an experimental campaign in 2018 in which dust was dropped vertically into the target chamber with a plasma beam flowing through it. The dust was composed of 5- and 9- μ m monodisperse W and was shaken through sieves before entering the chamber. The dust was initially dropped with no plasma present, and then a cylindrical plasma beam was generated with magnetic field values of 0.1, 0.2, 0.3, and 0.4*T*.

III. RESULTS AND ANALYSIS

RIDAT's machine learning algorithm was trained separately with JET, DIII-D, and M-PSI footage. The training datasets were randomly segmented from the image data, and the training set sizes were 365 tracks for JET, 597×10^3 for DIII-D, and 230×10^3 for M-PSI. The settings used for train-





(b)

FIG. 4. (a) A Calcam reconstruction of the plasma target chamber located in the Magnum-PSI machine [22]. (b) A typical still image taken by the Phantom visible fast imaging camera in Magnum-PSI. The illuminated section is the plasma beam.

ing and tracking were optimized in a number of preliminary runs. During these runs the change in brightness variable was removed from the machine learning feature set. This was because only the brightest and easiest to identify grains were used in training. Thus, a bias was created for brighter dust grains in training, which extended to a bias in tracking when the brightness feature was active. The settings with which RIDAT was run for each machine are shown in Table II.

A. JET

Each of the three trained algorithms were applied back to their corresponding machines to detect and track dust. For JET, over 30 dust tracks were identified over 34 different shots. An example track overlaid on a summed image from the divertor camera is shown in Fig. 5. From Fig. 5 it is clear to see that the RIDAT predicted track follows the bright streak of the physical dust grain well. This demonstrates RIDAT's ability to accurately and automatically track a dust grain across multiple

TABLE II. A table displaying the run settings for RIDAT applied to each plasma machine, where T_B is the threshold brightness, T_P is the threshold Baysian probability, Streak determines whether grains are treated as single grains or traveling streaks, and size is the number of datapoints used for training the machine learning classifier.

Machine	T_B	T_P	Streak	Size
JET	15	0.84	True	363
DIII-D	30	0.97	False	597×10^{3}
M-PSI	8	0.97	True	230×10^{3}

frames in video footage. However, the track shown in Fig. 5 is also incomplete, as the code has failed to track the first and last frames of the grain's path. This may be due to the significant variation in track direction at the start (top) and tail (bottom left) end of the path. This deviation in track features, coupled with the small (363 track) training data set for JET, could have been sufficient to incorrectly label the tail ends of the path as false. This issue may be resolved by using a larger, more varied training data set or by more carefully optimizing brightness and probability thresholds.

After the JET dust tracks were generated, the classification accuracy of the tracks was determined. This was done by overlaying detected tracks with camera footage and manually determining whether the track positions follow their real counterparts (similarly to the example in Fig. 5). The



FIG. 5. The pixel positions of a RIDAT-generated dust track from JET. Positions were overlaid onto a combined still from JET's NIR camera, found by summing the image data of every frame in this shot.



FIG. 6. The pixel positions of a random set of five RIDATgenerated tracks from shot 167345-11198 in DIII-D, overlaid on a still image from the same shot.

classification accuracy of RIDAT applied to JET was calculated to be 65%. This is lower than the value predicted by the Bayesian probability cutoff of 84%. The reason this classification accuracy is so low is most likely due to the quality of the JET footage, as the 35% of incorrect identifications were mostly camera noise or plasma emission. Additionally, the rapid deviation in track features mentioned earlier can be partially attributed to the JET camera's long integration time. Thus, for footage similar to the NIR camera in JET, RIDAT may be used as a first step in track generation; followed by more rigorous manual vetting.

B. DIII-D

RIDAT's application to DIII-D footage generated 925 dust tracks across two shots. Five example tracks are shown overlaid on an image from the DiMES camera in Fig. 6. The RIDAT generated dust tracks were again manually compared against camera data, and a classification accuracy of 99% was found. This high classification accuracy further demonstrates RIDAT's ability to generate high-quality track data from camera footage. Moreover, a large database of 925 tracks indicates RIDAT can successfully produce statistically significant datasets automatically. As such, it seems RIDAT is best suited to the type of data collected from DIII-D, in which there were large quantities of separated, bright, and consistent dust grains.

C. M-PSI

When applied to M-PSI, RIDAT produced over 1500 dust tracks across 18 shots. Five example tracks are shown in Fig. 7, overlaid on of a still image from the phantom camera. On manual inspection, the classification accuracy of the RIDAT tracks was determined to be 74%. This classification accuracy initially appears low. However, the majority of RI-DAT's inaccurate classifications for the M-PSI dust grains was because RIDAT could not distinguish between two very similar dust tracks. This was exacerbated by M-PSI's particularly bright background seen in Fig. 7. Thus, the performance of



FIG. 7. The pixel positions of a random set of five RIDATgenerated tracks from the 0.4T Plasma data set of Magnum-PSI, overlaid on a still image from the same camera. To distinguish the trajectories of overlapping tracks, a linear fit has been applied to each track.

RIDAT could be improved by more carefully selecting image processing parameters; parameters such as the threshold brightness and the number of frames over which a background is averaged.

IV. CONCLUSIONS

The RIDAT algorithm was designed and applied to camera footage from JET, DIII-D, and the linear plasma generator Magnum-PSI with varying levels of success. In JET, RIDAT detected over 30 dust tracks across 34 camera shots with a classification accuracy of 65% and incorrectly discounted sections of some correct tracks. In DIII-D, 925 dust tracks were generated from two camera shots, with an adequate classification accuracy of 99%. For Magnum-PSI over 1500 tracks were created with a classification accuracy of 74%. The issues of low classification accuracy and the misidentification of correct track sections highlight the drawbacks of RIDAT. However, these may be improved on by creating more robust training data, accounting for biases, and by more carefully selecting RIDAT parameters.

RIDAT could be used for a number of applications in the future. These include grain lifetime and splitting analysis, as well as 3D track reconstruction using dual stereoscopic camera systems such as those currently employed in the Mega Ampere Spherical Tokamak and TEXTOR [27,28]. For effective use, however, the code should be improved. Specifically by applying more optimized image processing techniques, such as those used in particle tracking for image microscopy [29]. In addition to significant algorithm improvements for the code, fine-tuning of the input parameters should also be performed for each future application. For example, increasing the algorithm's rolling frame window from three to four could result in more accurate classifications.

ACKNOWLEDGMENTS

We thank Michael Coppins for his expert supervision and advice that helped shape the direction for the RIDAT code. This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 and 2019-2020 under Grant No. 633053. We acknowledge the support of the Magnum-PSI Facility Team at DIFFER. The Magnum-PSI facility at DIFFER has been funded by the Netherlands Organisation for Scientific Research (NWO) and EURATOM. The views and opinions expressed herein do not necessarily reflect those of the European Commission. This work was supported in part by the U.S. Department of Energy under DE-FC02-04ER54698 and DE-FG02-07ER5491. This work was supported in part by the Engineering and Physical Sciences Research Council, under Grant No. EP/M001709/1. For the contribution to the M-PSI dust experiment, we thank Professor S. Ratynskaia [Space & Plasma Physics, School of Electrical Engineering Royal Institute of Technology (KTH) Stockholm, Sweden] and Marco De Angeli [Institute for Plasma Science and Technology (ISTP) Milan Italy].

- J. Contributors, E. Joffrin, T. Donné, G. J. van Rooij *et al.*, Nucl. Fusion **59**, 112021 (2019).
- [2] F. Sartori, G. de Tommasi, and F. Piccolo, IEEE Contr. Syst. Mag. 26, 64 (2006).
- [3] J. Lohr, DIII-D Research Operations Annual Report to the U.S. Department of Energy (1995).
- [4] J. D. Lawson, Proc. Phys. Soc. B 70, 6 (1957).
- [5] J. Roth, E. Tsitrone, T. Loarer, V. Philipps, S. Brezinsek, A. Loarte, G. F. Counsell, R. P. Doerner, K. Schmid, O. V. Ogorodnikova, and R. A. Causey, Plasma Phys. Control. Fus. 50, 103001 (2008).
- [6] T. Hirai, V. Barabash, F. Escourbiac, A. Durocher, L. Ferrand, V. Komarov, and M. Merola, Fusion Eng. Des. 125, 250 (2017).
- [7] A. Hassanein, G. Federici, I. Konkashbaev, A. Zhitlukhin, and V. Litunovsky, Fusion Eng. Des. 39, 201 (1998).

- [8] J. Winter, Phys. Plasmas 7, 3862 (2000).
- [9] A. Hassanein, V. Belan, I. Konkashbaev, L. Nikandrov, V. Safronov, A. Zhitlukhin, and V. Litunovsky, J. Nucl. Mater. 241, 288 (1997).
- [10] Summary of the ITER Final Design Report, ITER EDA Documentation Series No. 22 (International Atomic Energy Agency, Vienna, 2001).
- [11] J. C. Flanagan, M. Sertoli, M. Bacharis, G. F. Matthews, P. C. de Vries, A. Widdowson, I. H. Coffey, G. Arnoux, B. Sieglin, S. Brezinsek, J. W. Coenen, S. Marsen, T. Craciunescu, A. Murari, D. Harting, A. Cackett, and E. Hodille, Plasma Phys. Control. Fus. 57, 014037 (2014).
- [12] M. Sertoli, J. Flanagan, M. Bacharis, O. Kardaun, A. Jarvinen, G. Matthews, S. Brezinsek, D. Harting, A. Cackett, E. Hodille *et al.*, J. Nucl. Mater. **467**, 394 (2015).

- [13] L. Vignitchouk, P. Tolias, and S. Ratynskaia, Plasma Phys. Control. Fus. 56, 095005 (2014).
- [14] M. Bacharis, M. Coppins, and J. E. Allen, Phys. Rev. E 82, 026403 (2010).
- [15] G. Gervasini, E. Lazzaro, and A. Uccello, J. Fus. Energy 36, 25 (2017).
- [16] A. Y. Pigarov, S. I. Krasheninnikov, T. K. Soboleva, and T. D. Rognlien, Phys. Plasmas 12, 122508 (2005).
- [17] J. Martin, M. Bacharis, M. Coppins, G. Counsell, and J. Allen, Europhys. Lett. 83, 65001 (2008).
- [18] J. D. Martin, Ph.D. thesis, Imperial College London (2006).
- [19] C. Cowley and F. Patrick, cydcowley/RIDAT: Version 1.0.0 Release, https://github.com/cydcowley/RIDAT (2019).
- [20] A. Y. Ng and M. I. Jordan, in Advances in Neural Information Processing Systems, edited by T. G. Dietterich, S. Becker, and Z. Ghahramani (MIT Press, Cambridge, MA, 2002), pp. 841–848.
- [21] A. Huber, S. Brezinsek, P. Mertens, B. Schweer, G. Sergienko, A. Terra, G. Arnoux, N. Balshaw, M. Clever, T. Edlingdon *et al.*, Rev. Sci. Instrum. 83, 10D511 (2012).
- [22] S. Silburn, Calcam (2018).

- [23] M. Groth, M. Fenstermacher, C. Lasnier, R. Hernandez, J. Moller, and R. Sturz, Rev. Sci. Instrum. 74, 2064 (2003).
- [24] D. Rudakov, A. Litnovsky, W. West, J. Yu, J. Boedo, B. Bray, S. Brezinsek, N. Brooks, M. Fenstermacher, M. Groth *et al.*, Nucl. Fus. **49**, 085022 (2009).
- [25] E. Davis, J. Rost, M. Porkolab, A. Marinoni, and M. Van Zeeland, Rev. Sci. Instrum. 89, 10B106 (2018).
- [26] L. Simons, J. Holgate, C. K. Stavrou, B. Dudson, M. Thomas, P. Bryant, B. Harris, S. Ratynskaia, and T. Morgan, in *Proceedings of the 45th EPS Conference on Plasma Physics (EPS'18)* (European Physical Society, Prague, Czech Republic, 2018), pp. 1720–1723.
- [27] G. De Temmerman, M. Bacharis, J. Dowling, and S. Lisgo, Nucl. Fus. 50, 105012 (2010).
- [28] A. Shalpegin, L. Vignitchouk, I. Erofeev, F. Brochard, A. Litnovsky, S. Bozhenkov, I. Bykov, N. den Harder, and G. Sergienko, Plasma Phys. Control. Fus. 57, 125017 (2015).
- [29] N. Chenouard, I. Smal, F. De Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter *et al.*, Nat. Methods **11**, 281 (2014).