A novel image analysis technique for 2D characterization of overlapping needle-like crystals

DOI:
10.1016/j.powtec.2021.09.017

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Powder Technology

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
A novel image analysis technique for 2D characterization of overlapping needle-like crystals

Petros Neoptolemou a, Nishank Goyal a, Aurora J Cruz-Cabeza a, Anton A Kiss a, David J Milne b, Thomas Vetter a, b

a Department of Chemical Engineering and Analytical Science, University of Manchester, The Mill, Sackville Street, Manchester M1 3AL, United Kingdom
b Chemical Development, Pharmaceutical Technology & Development, Operations, AstraZeneca, Macclesfield SK10 2NA, United Kingdom

Abstract

Particle size and shape significantly affect powder processing and their end-product quality in a variety of industries. Imaging methods can successfully characterize populations of needle-like particles. Prior to off-line imaging, adjusting the particle density can reduce particle overlaps but increase measurement/processing times. Discarding data of overlapping particles, as most image processing algorithms do, biases particle size and shape distributions. Building on previous efforts, we here provide an image processing technique that accurately separates and sizes overlapping needle-like particles. Our algorithm combines edge detection, layer-stripping watershed segmentation and length approximation. To test the algorithm, a large number of real particle projections were randomly overlapped with various overlap intensities. Approximately 92-72% of the particles were detected and the particles’ dimensions were characterized with an accuracy of 87-75%, with these ranges corresponding to low and high overlap intensities. Overall, the algorithm removes biases to considerably improve characterization accuracy of powders containing needle-like particles.

Keywords: Particle separation, Needle-like particles, Image processing, Particle overlap, Image acquisition, Powder dispersion

1. Introduction

Particle size and shape distributions (PSSDs) in powder samples influence the properties of consumer products such as foods and pharmaceuticals [1] as well as their processing behaviour (i.e., thermal conductivity in packed beds [2], speed of filtration [3] and amount of liquid retained, flowability [4], tabletting [5]) and dissolution rate [6]. The need to control PSSDs has led to the establishment of a large number of in-process and off-line [7] characterization technologies such as the Focused Beam Reflectance Measurement (FBRM), Coulter Counter (CC), Laser Diffraction

*thv@lundbeck.com
**Current address: Department of Chemical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India.

Preprint submitted to Powder Technology September 2, 2021
as well as other various imaging techniques [8, 9] and custom-built laboratory setups [10, 11]. Some of those techniques yield only one-dimensional information of the particles (such as FBRM producing Chord Length Distributions) or make an assumption about the shape of the particles (e.g., LD produces diffraction patterns that are translated to particle sizes assuming particles are spherical); whilst a growing number of them are also capable of characterizing two dimensions (e.g., imaging). In parallel to the development of instrumentation, a significant number of scientists have also endeavoured themselves to the development of computational methods for the analysis and construction of PSSDs from acquired raw data [12–15] and to the development of mathematical models (based on population balances) to simulate the evolution of PSSDs from operational process parameters (e.g., in crystallization, granulation or drying processes).

Equant shaped particles can be characterized with various one-dimensional technologies and simulated readily with simple mathematical models [16]. Needle-like particles, however, pose various challenges both from an experimental and a computational viewpoint [17, 18]. Data from one-dimensional techniques have been used for the derivation of two-dimensional PSSDs by assuming constant aspect ratios or narrow size distributions of the particles’ width [12, 14, 19, 20]. For a highly accurate description of anisotropic morphologies, however, imaging provides the best solution since it allows for direct measurement of the length and width of particles [21, 22] and even for a reconstruction of three-dimensional shapes from dual projections [23–26].

The occurrence of overlapping particles during image acquisition is a well-known and common phenomenon resulting from either physical aggregation (e.g., during crystallization) or the overlap of object projections (objects laying in front of each other). The issue is found in a variety of scientific disciplines, such as the measurement of particulate products (among them crystals [27, 28]), biological cells [29] or even celestial objects [30]. In the Particle Engineering sector, quite often, the particles tend to take a needle-like form, instead of a spherical form as in the examples of planets/stars or cells. The majority of commercial software available for the processing of particle images and derivation of PSSDs simply discard images of overlapping needle-shaped particles based on characteristics of the measured projections (such as, for example, a reduced convexity). A few algorithms reported in the literature allow, however, to detect and account for aggregated particles [31] and characterize the total volume of the aggregates [32].

In order to derive accurate PSSDs of needle-like particles, correcting for particle overlaps is a necessity. Naturally, in a monodisperse distribution, the rejection of overlaps would not affect the average or the width of distribution: since all the particles have the same size, only one particle is, in principle, required to acquire the exact average. By contrast if the distribution is polydispersed (as in virtually all realistic cases), the rejection of overlapped objects causes a bias in the distribution. This is the case, because larger particles have larger areas of projection and therefore a higher likelihood to overlap in a given field of view than their smaller counterparts. Similar biases can occur when a mixture
of aspect ratios or particle shapes are present in a powder sample. The exact effect of overlap in the measurement is not part of this study, however simple simulations are available in the Supplementary Material, which prove that this effect is significant and that there is a necessity to address the overlap of needle-like particles. In off-line measurements, the number and extent of overlaps can be minimized by diluting the solution (for image acquisition in-suspension [34]) or by decreasing the sample amount whilst increasing the dispersion area (for dry measurements where the powder is dispersed on a plate under a microscope such as in this study). Minimising the overlaps in these ways, however, requires the use of extra material for the dilution and an additional number of recorded images so as to characterize a sample size large enough to be representative of the actual distribution (see SI and ref [35] for the effect of sample size on PSD properties). As a consequence of this, the image acquisition time and the image processing time increase. In addition, characterizing single particles in systems becomes challenging with increasing solid concentration, hence an image processing solution to deconvolute overlapped particles is essential.

The separation of touching particles has been widely studied and solved using watershed segmentation, a method particularly suited to address equidimensional particles (e.g. cube- and sphere-like particles [29, 30, 36–38]). The touching and overlap of needle-like particles, however, proves to be a difficult case for this technique alone due to the multiple nuclei (local maxima) detected in the Distance Transform (DT) for each particle (which often causes over-segmentation [39]).

In the last two decades, a few attempts to deal with the overlap of needle-like particles have emerged, which include methods of crystal edge detection, salient corner identification [28, 40] or deep learning [41]. The SHARC (Segmentation for High-Aspect-Ratio Crystals) algorithm [42] uses a slightly modified Burns line finder [43] to identify straight line features on a grayscale image. It then combines collinear line segments and detects the two edges of a crystal by identifying parallel line pairs [44]. Finally, the size of a crystal is defined by the properties of its two forming lines. The SHARC algorithm excels in crystal identification in noisy images, however its accuracy dramatically decreases with increasing overlap intensity. As mentioned by the authors, underestimation of the crystal size has been noted, part of it accredited to the algorithm recognising a segment of a crystal as a separate particle.

Ahmad et al [40] used a different approach, detecting corners from the grayscale image and identifying the salient corners from the binary edge image [45]. The algorithm showed great separation capabilities, but the requirement to detect at least three salient corners prohibits the identification of particles in certain overlap cases and makes detection of needle-like crystals challenging in cases of two corners missing as in the case of particle breakage.

In this paper, we present a separation method which specializes on separating overlapping needle-like particles. The algorithm makes use of the fact that the two straight edges of a crystal are more easily detectable than its corners, especially in conditions where the particles’ tips have been chipped off (this is a frequent problem for needle-like
crystals due to their tendency to break across their short dimension). An additional advantage of the proposed method is that it is based on the manipulation of binary images. The memory required to store binary images is 2-bit per pixel compared to typically 8-bit per pixel for grayscale images. As a consequence, the method allows for faster loading and saving times and requires less processing time as well as random access memory (RAM). After identification of parallel edges in an overlapped object, the algorithm initially calculates the edges’ line equations, extends the lines and retains only the pixels in-between. Then, using a layer-stripping algorithm [39] to identify markers for watershed segmentation, it separates touching needle-like particles. Next, the algorithm recognises if the type of overlap between a particle pair is partially-crossed and a length approximation is applied. In the final step, the length and width of the separated particles are measured by fitting an area conserved box to create the PSSD of the population. To further minimize biases on the PSSD, a simple image stitching method is utilized during the pre-processing step to increase the field of view. This allows particles that were cut off by the images boundary to be measured as well as increase the maximum length of particles that can be measured.

This article is organized in three different sections. Firstly, the methods are described which include a detailed illustration of the lab-built equipment used to acquire the images displayed, as well as the methodology of the algorithm including newly proposed parts and modifications to existing steps. Secondly, the algorithm is evaluated visually and statistically from results for a large population (≈19000 particles) of artificially overlapped L-glutamic crystals. Finally, errors and limitations of the algorithm are discussed.

We note, for clarity, that the term object is defined as the area of connected white pixels separated from the background in the pre-processing step. Hence in this study, 'object' could refer to both a single non-overlapped particle or a group of overlapped particles. We use the term subobject, thus, to refer to the resulting parts of an object separated after an image processing step. Finally, the terms particle and crystal are used interchangeably (the method works for any particles visible to the camera and in this paper the particles used are semi-transparent crystals).

2. Methodology

2.1. Equipment and experimental procedure

Three seeded slow cooling crystallization experiments in water were carried out to produce β L-glutamic acid crystals. The particles grown in this system tend to exhibit a needle-like habit [46], which is appropriate for the current research. The purpose of the four different treatments of the crystals was to combine the four populations to produce a wide PSSD, hence the algorithm could be tested on a large variety of particle sizes and aspect ratios. All three experiments started with a saturated solution at 50°C. The first batch was made by cooling down to 20°C with a cooling rate of 2°C hr⁻¹. To create the second batch, crystals from the first batch were temperature cycled. The
temperature was raised from 20°C to 35°C and cooled back down to 20°C four times with a cooling and heating rate of 2°C hr\(^{-1}\) and 5°C hr\(^{-1}\), respectively. The solution was left to reach equilibrium for 4 hours after each cooling stage and two hours after each heating stage. The third crystallization recipe was a combination of temperature cycling and wet-milling. Part of the suspension from the first batch was wet-milled at 20°C (see conditions below), then heated to 35°C and cooled down to 20°C with the heating/cooling rates and equilibration times equivalent to the second batch. This temperature cycle was carried out twice.

The wet milling was carried out using a rotor-stator IKA magic LAB (IKA-Werke, Germany) with an MKO module. The milling rotational speed was set at 5000 rpm for batch 3 and at 10000 rpm for batch 4. Both batches passed through the mill twice with a flow rate of 250 mL min\(^{-1}\). The crystallization and dissolution stages were carried out in a jacketed vessel.

The generated particles were then analysed with our imaging setup (see description below). Two-dimensional projections of non-overlapped crystals were randomly selected from all four batches and grouped into one crystal population, which was later used for the evaluation of the separating algorithm.

The imaging equipment used to acquire grayscale images of the L-glutamic acid crystals is shown in Figure 1. A Basler ace acA2440-20gm Monochrome GigE Camera (Basler AG, Germany) with 3.45 × 3.45 µm pixel size and 8.4 × 7.4 mm sensor size, was attached to a 4.0X SilverTL Telecentric Lens, which in contrast to a normal lens minimizes distortion [47]. The camera in combination with the 4x magnification lens allows each pixel to represent 0.8625 × 0.8625 µm of the real world and decreases the field of view to a size of 2.1 × 1.7 mm. The light source was placed on the opposite side and connected to a 60 mm Telecentric Backlight Illuminator, which allows only parallel light rays to pass through, hence increasing the contrast between particle boundaries and the background. The mentioned products and some of their fixtures were supplied by Edmund Optics Ltd, United Kingdom. Extra fixtures, such as the breadboard with a centered hole, were supplied by Thorlabs LTD, United Kingdom. The motorized stages were supplied by Zaber Technologies Inc, Canada. The xy-motorized stage (ASR100B120B-E03) has a maximum x- and y- movement of 120 × 100 mm with a step size of 0.15625 µm. The linear stage (X-LSM050-E) travels a maximum distance of 50.8 mm and is used to vertically move the camera to bring the objects on the glass plate in focus.

A powder disperser is necessary to ensure reproducible sample preparation as well as to give control over the number of objects per imaging area. A custom-designed and lab-built disperser was created from simple laboratory equipment, whose dispersion was acceptable to carry out this study. More complex powder dispersive set-ups have been designed elsewhere [48]. The diagram of the powder disperser is shown in Figure 2. A small sample amount (<0.1 g) is placed in the detachable sample container, which is connected to the tube inside the funnel. The funnel is sealed with a plastic stopper to prevent particles from escaping and a diaphragm valve is used as a push button to allow
pressurized air through to disperse the powder. The air pressure was decided by trial and error in a range of values between 1 to 5 barg. The pressure of the air used was set at 2 barg to prevent breakage of the crystals, while giving a good dispersion. In addition to the pressure, the number of particles dispersed also affect the number of overlaps since the area of the dispersion is fixed and equal to the area of the funnel’s opening. Hence, a trade off between overlap intensity and number of particles per frame exists. This interplay affects the image acquisition time, since a large enough sample size must be characterized to accurately represent a population’s size distribution (see supporting information). For example, dispersing a small number of particles would give better dispersion, but would require a larger number of frames to be obtained. The dispersion takes place on top of a plexiglas plate and the plate is then inserted onto the xy-motorized stage. The plexiglas plate and the detachable sample container of the disperser were constructed in our local workshop at the University of Manchester. Matlab 2019a was used for image acquisition and movement control of the stage, as well as the image processing presented in this paper.

2.2. Algorithm for separation of overlapped particles

By visually inspecting the dispersed crystals, three different groups of cases involving particle overlaps stand out as shown in Figure 3.

All three main overlapping cases are dealt with by using the algorithm described in the flowchart in Figure 4. The pre-processing step removes minor noise in the acquired images, converts the images to binary, detects individual objects and decides which ones are overlapped by calculating their convexity index (see equation 1). In Step 1,
Figure 2: Diagram of the custom-built powder disperser (left). Image of the dispersed crystals on the plate (right).

Figure 3: The three overlap cases of needle-like particles. The dashed lines represent the boundaries of the two separated crystals. The identification of the individual particles requires different processing steps in the different cases, as explained in the text.
straight lines are identified on the boundaries of the objects. Then in Step 2, the lines which are parallel are paired.

The parallel lines are extended and the image in-between is retained, while the pixels external to the parallel lines are cropped. In cases, where a subobject is detected more than once, then its smaller duplicate is discarded. Step 3 is used to separate any touching particles and Step 4 identifies the partially-crossed cases and applies a length approximation.

Finally, the separated particles can be sized using an area conserved box (ACB) and a two-dimensional particle size and shape distribution can be constructed. More details about the algorithm are given in the text below, together with the pre-existing Matlab commands used (given in Italics), that are also available in other widely used programming languages.

The pre-processing step starts by acquiring grayscale images as shown in Figure 5a, using our imaging set-up (Figure 1). The grayscale images are then binarized by calculating the global threshold, a value between 0 to 255, by using Otsu’s method [49], which can be applied by the imbinarize command. The complementary image is then calculated and minor noise, such as any thin scratches on the plate, is removed using an erosion followed by a dilation step. The imopen command was used for this purpose, where a layer with thickness of two pixels is removed from the object’s boundaries (erosion) and then added (dilation). Hence, small objects are completely removed and the remaining objects are returned to their original size. Subsequently, the voids in the crystals are filled in (imfill), as shown in Figure 5b. The boundaries of each object are found using the command bwboundaries, which identifies separate objects to be the individual areas with connected white pixels. The object’s boundaries are shown in Figure 5c. Setting a convexity threshold, $T_c$, is a simple condition used to separate single from overlapped particles. The convexity index is calculated as:

$$f_c = \frac{A_p}{A_{ch}}$$

where, $A_p$ is the area of the object and $A_{ch}$ is the area of the object’s convex hull. The area of the object is simply the sum of the white pixels, while the area of the convex hull is calculated as the sum of the white pixels produced by the property ‘ConvexImage’ of the regionprops command. The boundaries of an example of the convex hull of a particle are shown in Figure 5d.

To calculate the optimum convexity threshold, $T_c$, 500 objects were selected so as to cover a large range of sizes, aspect ratios and have about an equal amount of both overlapped and non-overlapped cases. The examples were manually classified as overlapped or non-overlapped. The convexity threshold value was varied until the classification error was minimised, i.e., reached the maximum number of correctly classified objects. The optimum convexity threshold was determined to be 0.865 (see Supplementary Material).

In Step 1, the Hough Transform Method is utilized to extract straight lines from the boundaries of objects
Figure 4: The flowchart presents the four simple steps required to separate needle-like particles in touching cases, partially- or fully- crossed overlaps.
Figure 5: Pre-processing of images. a: Grayscale image of needle-like crystals acquired using our imaging setup. b: The image is binarized using the global threshold calculated using Otsu’s method. c: The individual objects are detected and their boundaries are highlighted. The object in the black dashed rectangle is selected for further processing. d: A fully-crossed overlap with a convexity value of 0.35.
constructed from overlapping particles, as shown in Figure 6a. The following equation gives the normal form [50] of the lines, which is used to create the Hough Space by applying the `hough` command in Matlab:

\[
\rho = x_L \cos(\theta_L) + y_L \sin(\theta_L)
\]  

(2)

where \(\rho\) is the perpendicular distance from the line to the origin, \((x_L, y_L)\) are the Cartesian coordinates of the points that lie on the line and \(\theta_L\) is the angle of the line. Applying the `houghpeaks` command, the two-dimensional Hough Space is discretized using bins of size 0.5 pixels and 0.5 degrees angle across the \(\rho\) and \(\theta_L\) dimensions, respectively. Hence, the number of bins across the \(\theta_L\) dimension is constant at 360, but the number of bins of the \(\rho\) dimension is proportional to the difference between the largest and smallest distance from the origin (different for each object). The extracted peaks are a combination of line segments around each peak in the Hough Space, which have a similarity of 1% of the maximum size of each of the \(\rho\) and \(\theta_L\) dimensions. Finally, lines are formed from the peaks using the `houghlines` command. The considered line segments have a minimum length of 10 pixels and lines are formed from points that have less than 8 pixels distance in-between to prevent forming lines across the width of thin crystals.

All the straight lines are compared to each other to identify groups of line segments that could be further merged into one. The method used [51] takes pairs of lines that have a maximum angle difference, \(\Delta\theta_{MAX}\), and creates a new merged line that passes through the centroid, defined by the four endpoints of the line pair, and has a length-weighted angle.
The centroid of two lines is defined as:

\[ x_G = \frac{l_i(a_x + b_x) + l_j(c_x + d_x)}{2(l_i + l_j)} \]  

(3)

\[ y_G = \frac{l_i(a_y + b_y) + l_j(c_y + d_y)}{2(l_i + l_j)} \]  

(4)

where \( l_i \) is the length of the segment with endpoints \((a, b)\) and \( l_j \) is the length of the segment with endpoints \((c, d)\).

The length-weighted angle is calculated by:

\[ \theta_r = \frac{l_i \theta_i + l_j \theta_j}{l_i + l_j}, \text{ for } |\theta_i - \theta_j| \leq \frac{\pi}{2} \]  

(5)

\[ \theta_r = \frac{l_i \theta_i + l_j \left( \theta_j - \frac{\pi}{2} \right)}{l_i + l_j}, \text{ for } |\theta_i - \theta_j| > \frac{\pi}{2} \]  

(6)

where \( \theta_k \) is the angle of line segment \( k \).

In the next step we define a frame with axes \((X_G, Y_G)\), as shown in Figure 7, where the \( X_G \) axis is parallel to the merged line with angle \( \theta_r \). We use a translation followed by a rotation to transform the endpoints, \((\delta_x, \delta_y)\) to \((\delta x_G, \delta y_G)\), which are their corresponding coordinates on the \((X_G, Y_G)\) frame:

![Figure 7: Merging of two line segments. The figure was adapted based on Fig. 7 from reference [51].](image-url)
\[ \delta x_G = (\delta_x - y_G) \sin(\theta_y) + (\delta_x - x_G) \cos(\theta_y) \] (7)

\[ \delta y_G = (\delta_y - y_G) \cos(\theta_y) - (\delta_x - x_G) \sin(\theta_y) \] (8)

In addition to the difference in angle, two more criteria are used to test if two segments are part of the same line:

\[ l_r = \|a_{X_G} - b_{X_G}\| - \|c_{X_G} - d_{X_G}\| \leq d_{\text{MAX}_{X_G}} \] (9)

\[ \|k_{Y_G} - p_{Y_G}\| \leq d_{\text{MAX}_{Y_G}} \] (10)

where \( l_r \) is the length of the merged line defined by the furthest endpoint projections on the \( X_G \) axis and \( d_{\text{MAX}_{X_G}} \) is the furthest distance between two lines across the \( X_G \) axis allowed for merging. Also, \( k_{Y_G} \) and \( p_{Y_G} \) are the most distant endpoints across the \( Y_G \) axis and \( d_{\text{MAX}_{Y_G}} \) is the distance between the largest endpoint projections on the \( Y_G \) allowed for merging. We set \( d_{\text{MAX}_{X_G}} \) equal to 50 pixels as a sufficient number to connect collinear lines and we chose a small value of 8 pixels for \( d_{\text{MAX}_{Y_G}} \), hence avoiding merging the edges of thin crystals. Figure 6b shows the merged lines.

To identify the pairs of lines that are potentially the edges across the width of the crystals, we use the \( \Delta \theta_{\text{MAX}} \) with the same value used for line merging and set a minimum length of endpoint projections. Hence, equations 3-8 are reused, however in this condition, the second pairing criterion is defined as:

\[ \|k_{Y_G} - p_{Y_G}\| > d_{\text{MIN}_{Y_G}} \] (11)

where \( d_{\text{MIN}_{Y_G}} \) is a user defined parameter referring to the minimum distance along the \( Y_G \) axis allowed for lines to be paired and has the same value as \( d_{\text{MAX}_{Y_G}} \). Furthermore, all the possible combinations of pairs of lines are calculated and the pixels in-between the line pairs are retained, as shown in Figure 8. The parameters we have chosen for the line extraction and line merging are flexible enough to allow identification of lines on a wide range of crystal widths. However, in some cases, this causes multiple pairs of lines to enclose the same crystal, as seen in image a-b and c-d of Figure 8. Another issue is that part of the crystal could be enclosed in-between two parallel lines as seen in Figure 8 image e.

To choose the correct particles, the subobjects are sorted in order of descending area. Then, each subobject is overlapped with the subobjects that have lower area, iteratively. If the area overlap, \( A_O \), is larger than 90% of the smaller crystal’s area, then the smaller crystal is rejected/discarded. As seen in Figure 8, subobject b is rejected by a and subobject d is rejected by c. The last subobject e was formed from two almost collinear lines that did not fulfil
Figure 8: Step 2 - Reject/discard duplicate subobjects. The retained subobjects in-between all the parallel line pairs are identified and are shown in descending order with respect to their area (left). The crosses show the duplicate particles that are discarded. A duplicate is defined as the subobject that at least 90% of its area overlaps with a larger subobject. The ticked particles (left) are the accepted subobjects (right).

the merging criteria but barely fulfilled the criteria to be considered a parallel pair. Subobject e overlaps with a and is discarded.

Figure 9: The resulting subobjects after the application of image pre-processing and the first two steps on a partially crossed case (Top row) and a touching case (Bottom row). The particles a-d are the result after cropping between the parallel line pairs and discarding duplicate subobjects. The last column indicates the overlapping area between the subobjects.

Up to this point, the method separates almost perfectly fully-crossed overlaps. However, the partially-crossed and touching cases still need to be addressed. As Figure 9 shows, Steps 1 and 2 overestimate the size of one of the crystals in a partially-crossed (top) overlap, in this case a small part of particle a causes the length overestimation of
b. Similarly, the length of one of the crystals (crystal c) is overestimated in a touching case (bottom row). The error is better visualized when the image area included in both particles is highlighted in white (right most panels in Figure 9).

The processing continues with Step 3, where watershed segmentation is used to correctly separate any touching crystals. The first step is calculating the Distance Transform (DT) of the acquired binary subobjects, to obtain the smallest distance from each pixel of the subobject to its boundary, as shown in Figure 10a. Usually, the next step is to find the local maxima and obtain a marker (mask) for each of the touching particles. Needle-like particles, however, tend to have multiple local maxima and watershed can cause them to oversegment. Several methods have been proposed to prevent over-segmentation [52, 53], of which a layer-stripping technique [39], with small modifications, showed satisfying accuracy of marker identification for the particles used in this work. The DT can be normalized and discretized to obtain 10 layers. The main step is to start removing boundary pixels layer by layer. After the removal of each subsequent layer, the convexity of the subobject/s is calculated. If one of the subobjects exceeds the pre-defined convexity threshold then the object is removed from the array and saved as a mask. The DT of the object in Figure 10a, shows clearly that removing only one layer gives two masks, shown in Figure 10b. In all touching cases tested, the two new masks have higher convexity than the object input. Hence, the threshold that indicates if a stripped subobject can be used as a mask, must always be higher than the threshold of the object prior to stripping, as well as, higher than $T_c$ to ensure it is a convex particle. Being unable to classify which subobjects require watershed segmentation (touching cases), forces all of the subobjects to go through this process. Therefore, the layer-stripping process is terminated when the 7th layer is removed giving us the ability to separate touching objects connected by a neck with thickness up to 7 layers while preventing single particles with multiple nuclei to over-segment. This restriction leaves the crystals from Step 2 in Figure 8 and crystals a, b and d in Figure 9 intact, since no further segmentation is required.

Up to this point, the edges of the crystals in each of the three overlap cases have been successfully identified. The fully-crossed and touching cases have been resolved, however the length of the partially-crossed particle b shown in Figure 9 is still overestimated. This length error arises due to part of the area of particle a (Figure 9) being counted in particle b’s area. To solve this problem an approximation of the length can be done. Step 4 starts by identifying which subobjects are partially-crossed by removing the overlapped area between pairs of overlapped particles. As seen in Figure 11 each of the three overlapped cases yields a different number of subobjects. The case where three subobjects are observed is identified as the partially-crossed overlap. In this case, the particle that was not divided in two, enclosed in green boundaries, is the one that the length correction needs to be applied on. The cases of fully-crossed overlaps or touching overlaps on the other hand, yield four and two subobjects, respectively, and require no further processing.
Figure 10: Step 3 - Layer stripping and watershed segmentation. 

**a.** The Distance Transform of the object is calculated and discretized to 10 layers. 
**b.** The boundary pixels are removed layer by layer until the stripped objects have a higher convexity than the starting object or until layer 7 is stripped. In this example only one layer was removed and two convex markers were obtained. 
**c.** The watershed segmentation separates the original object into two, where the crossed subobject is classified as a duplicate and discarded because more than 90% of its area overlaps with particle d in Figure 9.

The length correction (or cut), $L_C$, is the distance that needs to be removed from the overestimated particle. Figure 12 shows the case of two partially-crossed particles where the length of particle 2 is overestimated. The cut is defined as the line (dashed line) passing through the crossing point (red dot) of the particles’ axes (black solid lines) and is perpendicular to the particle’s axis whose length is overestimated (particle 2). The line equations of the particles’ axes can be simply calculated from their centroid, shown as $(x_i, y_i)$, where $i$ is the particle’s number; and the gradient of the particle’s axis, $m_i$, is the length weighted average of the gradient of its two parallel lines, calculated by Equations 5 or 6. The length correction can be directly calculated by:

$$L_C = L_o - L_a$$  \hspace{1cm} (12)\]

where, $L_o$ is the overestimated length and $L_a$ is the approximated length of the particle. Hence, the potential error can be given as $\epsilon_p = \frac{L_C}{L_a}$, and a maximum threshold, $T_{\epsilon_p}$, can be set to discard particles that the approximation could give a large overestimation/underestimation, i.e., $\epsilon_p \leq T_{\epsilon_p}$. In this study, $T_{\epsilon_p}$ was arbitrarily set at 0.5.
2.3. Image stitching

A simple image stitching method was implemented to remove any biases caused by rejecting particles that are cut off by the boundary of the image. Due to the limited field of view large particles have a higher likelihood to end up on the edge of an image than their smaller counterparts. Evidently, particles larger than the field of view could not be sized at all.

Controlling the XY-stage, the plate can be accurately moved along the x- and y-axes. The manufacturer reports a maximum deviation in the actual position of the stage of 2 μm over a 10 mm movement and a maximum error of 40 μm, which can occur when moving over the entire length. In addition, the movement error increases by vibrations caused by handling the imaging setup as well as other equipment in the lab. Both errors are accounted for in the algorithm. The image stitching is implemented after binarization is applied, i.e., at the beginning of the pre-processing step. To correct for the movement error, consecutive frames are acquired with an overlap between them. Figure 13, shows a representation of two consecutive frames that were acquired along the x-axis. The two yellow dashed lines indicate the start of Figure 13b inside 13a and the end of Figure 13a inside 13b, respectively (i.e., the overlap).

The first step is to identify two features from Figure 13a that also exist in Figure 13b. Ideally two such features

Figure 11: Step 4 - Is the subobject partially-crossed? Top row: Resulting subobjects after steps 1-3 were applied. Bottom row: Classification in the three overlap categories. Each category gives a different number of subobjects when the overlapped area is removed. The case that gives three subobjects is the partially-crossed case and the particle that remained undivided enclosed by green boundaries is the one that requires a length correction.
should lie in the overlapped area. However, due to the movement error, the two features identified must be away from
the overlap area’s boundaries, i.e., inside the scanning area shown in Figure 13a. A feature is defined as an area with
size of 20x20 pixels, in which the percentage of white pixels is between 20% and 80%. This range was decided by
trial and error. The lower limit ensures that the feature is not entirely black and it is large enough to be unique, while
the higher limit ensures that a feature is not entirely made up of white pixels. The algorithm scans for one feature
from the left-top to the right-bottom of the scanning area, while for the other feature it scans from the right-bottom to
left-top, to ensure that the two features (red solid squares) selected are not identical. The scanning stops as soon as
any two areas meet the mentioned criteria.

The second step involves scanning the overlapped area of Figure 13b to locate the areas that match the features
chosen from Figure 13b. To take advantage of the moving accuracy of the stage the scanning area is limited to a small
square around the expected (ideal) position of each feature. Acquired images showed that the actual position deviation
in-between consecutive images is up to 10 pixels hence, to ensure this error is accounted for an area of size 31x31
pixels is scanned. The centre of the scanning area is the ideal position of each feature giving a possibility of up to 15
pixels of deviations.

To identify the matching features, two areas from Figure 13b are subtracted from the two features from Figure
13a. The sum of the absolute values of the resulting two matrices is the comparison metric. The new criterion is to
compare areas which have exactly the same distance in the x- and y-axes as the two features, shown as the feature

![Diagram](image-url)
Features

Area with high metric value

Area with zero metric value

Overlapped area boundary

Feature distance

Scanning area

distance. When all the possible comparisons are done, the position of the features is matched to the position of the areas that gave the lowest comparison metric value. Figure 13b shows an example where the comparison metric had a large value (dissimilar), in dotted squares, and the position of the areas where the metric’s value is expected to be zero (similar), in dash-dotted squares. Similarly, the images are stitched from top to bottom.

Initially the first column of images is stitched (top to bottom) and then each row follows, stitching images to their left neighbour. A $3 \times 3$ matrix of stitched images is created. The stage moves 1.2mm between consecutive shots which corresponds to overlap distances of 1057 pixels and 657 pixels in the x- and y-directions, respectively. The reason of this value is to select a large enough overlap area to ensure that particles lie inside. The image stitching process allows the field of view to increase from $2.1 \times 1.7$ mm to $4.5 \times 4.2$ mm, which is well above the maximum size of the particles in the applications considered. However, in principle, the field of view could be further extended if required. Rarely, only one or no features are identified in which case the images are stitched at their ideal position.

2.4. Particle sizing using area conserved box

Two of the most common ways to measure the dimensions of elongated particles are a) calculating the minimum and maximum Feret diameter (or caliper diameter), which would correspond to the length and width of the particle, respectively and b) calculating the minimum area bounding box [54], where the longer and shorter dimensions would
correspond to the length and width of the particle, respectively. Using the Feret’s minimum and maximum diameters causes the characterization of the aspect ratio to drop for rectangular objects, since their diagonal is considered the length. However, using the minimum and maximum Feret’s diameters the equivalent dimensions of a rectangle can be calculated [55]. Fortunately, there is no need to know how rectangular the object is or post-process the calculated dimensions when using the minimum area bounding box (MABB). However, the MABB could give an overestimation in case the particle has an abnormality on one of the two faces, such an example is given in Figure 14b, where the width, \( W_i \), is overestimated due to an outgrowth. To solve the issue, an area conserved box (ACB) can be used to fit a rectangle on the particle’s projection that avoids any overestimation due to surface abnormalities.

![Figure 14: Comparison between minimum area bounding box and area conserved box. Image a: the grayscale image of the crystal. Image b: fitted minimum area bounding box (MABB). Image c: The dimensions used as initial guesses of the ACB optimization. Image d: The area conserved box (ACB), which arises from maximising the area overlap between the box and the particle.]

The length \( L_f \), width \( W_f \), position \((x_f, y_f)\) and angle \( \theta_f \) are set as the unknown parameters that make up the ACB of a particle (Figure 14d). These parameters are adjusted until the area of the box covers as much area of the particle’s projection as possible. The negative value of the area overlap, \( A_o \), can be minimized using the Matlab function \textit{fminsearch}, where the initial guesses of the parameters are shown in Figure 14c. The initial guess of the angle, \( \theta_g \), is set as the angle of the MABB. Then, the centre position guess \((x_g, y_g)\) is set as the barycentre of the particle’s projection. Finally, as the name suggests, the area of the box must be fixed at the same value of the particle’s area. Hence, if the aspect ratio of the initial guess box is set to be equal to the aspect ratio of the MABB, i.e., \( L_i/W_i = L_g/W_g \), then the first guess of length is calculated by:

\[
L_g = \sqrt{A_o \frac{L_i}{W_i}}
\]
where \( A_p \) is the known particle area, \( L_i \) and \( W_i \) are the length and width of the minimum area bounding box. Then, the width guess is simply given by:

\[
W_g = \frac{A_p}{L_g}
\]  

Since the width is inversely proportional to the length it is a dependent parameter and the function to be minimized is:

\[
f(\theta, L, x, y) = \frac{A_o(\theta, L, x, y)}{A_p}
\]

with a division by the particle’s area to give a fraction that is independent of the size of the particle. The command `fminsearch` finds a local minimum of arbitrary non-smooth functions such as \( f(\theta, L, x, y) \). Since a good initial guess is available (by the minimum bounding box), the identified local minimum was always found to be an acceptable result in the cases investigated here. The optimizer is terminated when two subsequent iterations yield function values with a difference less than 0.005, i.e. when the difference of area overlap between the last two iterations is less than 0.5%. As seen in Figure 14d, the ACB, in an indirect way, removes the width overestimation caused by the particle’s outgrowth.

In order to make the calculation of the ACB more efficient computationally, the position of its centre can be kept the same as the initial guess \((x_g, y_g)\). A test on 270 randomly selected particles, showed that excluding the centre from the optimization, results in an absolute error less than 0.2% on the volume-weighted mean, as well as, an error less than 0.8% on the volume-weighted standard deviation, for both length and width. In addition, the corresponding number-weighted properties showed less than 0.4% error. The time required to calculate the ACB was decreased approximately by half. One can therefore conclude that the exclusion of the box’s centre from the optimization gives an insignificant error compared to the advantage of saving half the processing time.

2.5. Artificial overlap for accuracy evaluation

Evaluation of algorithms for the separation of needle-like particles has been presented before, by comparing manual size characterization to the size characterization of the separation algorithms [40, 42, 56]. However, manual characterization presents the issue of inconsistent human perception, hence each individual can identify a different number of particles in each image with an increased deviation with increasing overlap intensity. To overcome the problem in this paper and avoid labour intensive characterization from multiple people, artificially overlapped populations were created. First, about 19000 single crystals of L-glutamic acid were chosen. Firstly, the images were pre-processed as discussed in Section 2.2. Secondly, all of the particles were overlapped to create a population containing only overlapping pairs and no single particles. Finally, the particle dataset was re-overlapped to create
two more populations which contain either only trios or only quintets (five overlapped particles). In a real case scenario, single (non-overlapped) particles would be present. However, in the case of evaluating the performance of the separation between particles, but not the separation of particles from the background, the number of single particles included would impact the accuracy of the test. Since single particles are always characterized accurately as long as they are correctly separated from the background, including any singles would result in a bias favouring the accuracy of the algorithm. Hence, we did not include any singles in the evaluation tests. The population containing solely quintets had the purpose of testing the algorithm at high solid concentration or high aggregation scenarios where a population consists of chunks made of several particles.

Figures 15a to 15c show three randomly selected single particles, which are then binarized as shown in Figures 15d to 15f. Then, the single particles are characterized and the binary images are randomly orientated and randomly positioned in a new image. The space that contains the crystals was made two times larger than the diagonal of the minimum area bounding box of the largest crystal in the group, to make sure there is enough space for all the crystals to be placed. The new image is accepted only when one object is detected, meaning that all the crystals have overlapped, and when the object’s convexity is lower than the threshold $T_c$. The particles are reorientated and repositioned, until the two criteria are met.

The proposed algorithm for separation is run for each overlapped group and the detected crystals were compared to the original ones. The similarity between each detected and each original particle in a group is decided by their size and angle. All detected particles are matched to an original, except from one case: when the number of particles identified is higher than the existing number of particles (e.g. four particles identified in a trio) then the extra detected particles cannot be matched. While in the case where less particles are identified than the existing ones, all of the detected particles are forced to find a match from the correct population (the most similar). The PSSDs of the original (referred to as correct) population and the detected populations (referred to as approximated) were compared visually and in terms of their number-weighted average and standard deviation. For any population the number-weighted standard deviation is given by:

$$\sigma_D = \sqrt{\frac{\sum_{i}^{N_T} (x_{D,i} - \mu_D)^2}{N_T}}$$

(16)

where $D$ represents a dimension, i.e., either length $L$, or width $W$, $\mu_D$ is the mean value of dimension $D$, $N_T$ is the total number of particles in the population and $x_{D,i}$ is the size of dimension $D$ of particle $i$.

Finally, the absolute value of the relative error between the correct and the approximated projection of particle $i$ is given by:

$$\epsilon_D = \left| \frac{x_{D,\text{c}} - x_{D,\text{a}}}{x_{D,\text{c}}} \right| \times 100$$

(17)
Figure 15: Artificial overlap of a trio. Three single crystals from an L-glutamic acid population are randomly chosen (a-c). After image pre-
processing (d-f), the binary single particles are randomly orientated and positioned in a new image (g).
where \( x_{D,i} \) is the size of dimension \( D \) of particle \( i \) from the approximated population \( a \). \( x_{D,i} \) has similar meaning, but refers to the correct population \( c \). It should be noted that only approximated particle projections which have been matched to their correct projection can be compared (as explained above).

3. Results and Discussion

3.1. Visual evaluation of “as-acquired” particle overlaps

Visual observation of the algorithm’s separation results allowed an initial evaluation of its performance. In Figure 16, the separation of overlapped crystals is shown at different overlap intensities, increasing from left to right. Figures 16a to 16c show the acquired grayscale image for each case and 16g to 16i show the resulting ACB fit for each separated particle. The Figures 16d to 16f show the intermediate pre-processing step, which is the binarized complementary projection of each object. Figure 16g shows the separation of two partially-crossed crystals. Figure 16h is a trio overlap, where crystal 1 is fully-crossed with crystal 2 and crystal 1 is partially-crossed with crystal 3. The higher intensity case in Figure 16i, shows the separation of four overlapped crystals, where particles (1-2) are touching, (2-4) and (3-4) are partially-crossed and (2-3) is a fully-crossed case. In conclusion, the visual evaluation of the algorithm in these common overlapped cases shows great promise in accurately separating overlapped needle-like particles. These images were cropped from the actual 3×3 stitched image as it is too large to present. In Figure 16g some objects at the top are not recognized because they touch the image top boundary, this phenomenon still exists but was minimized by applying image stitching.

For the sake of testing the algorithm’s limits, extreme overlapping cases were studied. In Figure 17, multiple crystals are overlapped in a confined space resulting in a high particle density (particles per area). Such overlaps are rarely observed with the imaging equipment and dispersion method presented in Section 2.1, but the algorithm’s good separation performance indicates its potential to be effective in systems that undergo a high degree of agglomeration or in in-process single projection imaging techniques for high solid concentration systems. Similar to Figure 16 the images in the top row of Figure 17 are the grayscale images, while the two in the middle row are the pre-processed images and the bottom row shows the ACB fits on each separated crystal. In Figure 17e an obvious error occurs with crystal 1, where due to cropping all the pixels in-between the two parallel edges of crystal 1, two (or three) crystals are considered as one, and its convexity index is higher than the threshold \( T_c \), hence considered a single particle. However, the ACB fit gives a smaller error, than what a minimum area bounding box would. In Figure 17f crystal 1 is clearly overlapped, however two straight lines were not identified on the other crystal because the straight part of its edges is not exposed. The second object of interest in Figure 17f is another case of overlap, which could be considered an extreme case of aggregation as well. The first error occurs between crystals 2 and 3, where the space in
Figure 16: Visual results of the separating algorithm. Images a to c show the grayscale data, where the overlap intensity increases from left to right. Images d to f show the resulting binary projections from the image pre-processing step. Images g to i show the ACB fit on the separated crystals.
between the touching crystals is filled during the image-preprocessing, shown in d, and instead the pair is classified as partially-crossed, while the length of crystal 3 is overestimated. Furthermore, regardless if the pre-processing step works well on filling the bright areas inside semi-transparent crystals, when they create a closed loop, the inside empty area will also be filled. In such cases, non-existent particles can be identified when parallel lines are paired from neighbouring actual crystals, as in the case of crystal 4. Nevertheless, even in these extreme cases of overlaps, the proposed algorithm seems to perform quite well.

3.2. Artificial overlap evaluation

Four cooling crystallization experiments with different treatment of the crystals were used to obtain as many sizes and aspect ratios possible. The crystals were randomly selected from these four sources and combined into one big dataset. Three populations of artificially overlapped objects were created from images of these crystals as described in Section 2.5. Each of the populations consisted of either pairs, trios or quintets, in order to assess the accuracy of the separating algorithm in various overlap intensities. The actual normalized two-dimensional PSSD is presented as a contour plot in red in Figure 18 in the top row. The levels of the contours are at 0.2, 0.4, 0.7 and 0.9 of the maximum of the distribution, moving from pale to darker red. The detected PSSD was constructed from the crystals separated using the proposed algorithm and is shown in blue lines at the same contour levels as the correct PSSD, increasing level while moving from the outside to the centre. Each PSSD includes an image illustrating the overlap intensity of the corresponding population and the identified boundaries of each separated crystal in dashed lines. Observing Figure 18, the match of the PSSDs is satisfying, however it shows that the error of the deviation of length increases with increasing overlap intensity and it becomes obvious with the extremely overlapped population in 18c. On the other hand, plotting the cumulative size distribution (bottom row), a higher amount of small crystals is detected than the correct distribution. Both of the mentioned observations show a widening of the correct distribution, especially in length. This mainly occurs due to the underestimation and overestimation of the particle length in Step 4, but can also be affected by cases of under- and over-segmentation. More detailed information in Table 1 shows that there is a slightly increasing error in the length average, $\mu_L$, with increasing overlap intensity, as expected from observing the PSSDs in Figure 18. An additional comparison is made with the PSSD properties of correct, matched and detected populations. The presented information shows that there is an insignificant difference between the PSSD averages of the matched and detected particles, and the overestimation of length arises from non-detection or rejection of smaller crystals of the original population. In addition to the reasons provided from the visual evaluation, the average size overestimation can be attributed to the algorithm that carries out the artificial overlap. Since it does not prevent any crystals from completely hiding others, smaller crystals have a higher probability to be completely hidden, hence not being measured. This causes a bias in the size averages towards larger particles. Moreover, Table 1, shows that
Figure 17: Visual results of the separating algorithm on extreme cases. Images a and b show the grayscale image acquired of extreme overlapped cases. Images c and d show the resulting binary projections from the image pre-processing step. Images e and f show the ACB fit on the separated crystals.
Table 1: Particle population properties of the correct, matched and detected particles.

<table>
<thead>
<tr>
<th></th>
<th>Pairs</th>
<th>Trios</th>
<th>Quintets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>Matched</td>
<td>Detected</td>
</tr>
<tr>
<td>$\mu_L$, $\mu m$</td>
<td>271</td>
<td>280</td>
<td>271</td>
</tr>
<tr>
<td>$\mu_W$, $\mu m$</td>
<td>30</td>
<td>31</td>
<td>29</td>
</tr>
<tr>
<td>$\sigma_L$, $\mu m$</td>
<td>157</td>
<td>159</td>
<td>162</td>
</tr>
<tr>
<td>$\sigma_W$, $\mu m$</td>
<td>18</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>$N_T$</td>
<td>18744</td>
<td>16703</td>
<td>17173</td>
</tr>
</tbody>
</table>

the standard deviations of length, $\sigma_L$, and width, $\sigma_W$ (Equation 16), show a positive error from both the correct and matched populations, which is an effect of the underestimation and overestimation of the particles’ size, resulting in a slightly larger variety of sizes. In addition, the number of detected particles decreases with increasing overlap intensity. From about 19000 particles the amount of crystals detected were 91.6% in pairs, 81.0% in trios and 71.9% in quintets. This decrease mainly occurs due to the non-identification of the relatively smaller particles, which in some cases are completely hidden or are discarded from Step 4 if their length approximation has a maximum potential error larger than the error threshold, $T_{sp}$.

Figure 18: Particle size and shape distribution (PSSD) comparison between the actual L-glutamic acid population (Correct) and the population of the crystals that were detected from the proposed algorithm (Detected). The three contour plots (top row) of the two-dimensional distributions are in order of overlap intensity. The artificial populations include either overlapped a. pairs, b. trios, or c. quintets. The contour levels are drawn at 0.2, 0.4, 0.7 and 0.9 of the maximum of the distribution, from pale to darker red for the actual PSSD and from the outside to the centre for the estimated PSSD. Each plot includes an image example of the overlap intensity it corresponds to, as well as, the boundaries of the separated crystals in dashed lines. The bottom row shows the cumulative distributions of length and width for each of the three populations.

We can also compare the size of the detected crystals to their original (i.e., matched crystals) one by one, for each
of the three populations, as shown in Figure 19. Here, the cumulative fraction of particles with a relative error in length (top) or width (bottom) below a given value is shown. For instance, the top figure shows that for the case of the overlapped pairs roughly 30% of the particles are identified with less than 1% error in their length, about 90% of the particles are identified with less or equal than 10% error, etc. The calculation of length, as shown becomes more challenging with increasing overlap intensity, while, the calculation of width is much less sensitive to overlap intensity. The increase in the absolute value of the relative length error is expected, since with increasing overlap intensity the probability that the tips of a crystal are hidden increases, hence more partially-crossed cases exist. This increases the number of particles that have their length approximated in Step 4. To decrease the amount of particles with length overestimation, the maximum error threshold, $T_{\Delta L}$, could be decreased, but this would also decrease the number of particles identified. Similarly, the straight edges which define the crystal’s width have an increasing probability to be hidden with increasing overlap intensity. However, if not detected, then the crystal is simply not matched and not included in this error evaluation, justifying why the overlap intensity shows almost no effect on the accuracy of the width’s calculation. As shown in Figure 19, the separating algorithm shows high sizing accuracy when comparing detected particles with their original (matched). In more detail, 86.7% of the particles in pairs, 83.4% in trios and 74.6% in quintets have less than 10% absolute value of relative error in length. Similarly, the percentage of particles with less than 10% absolute value of relative error in width is 78.6% in pairs, 79.3% in trios and 77.7% in quintets.
4. Conclusions

An algorithm to separate overlapped needle-like particles was created by combining edge detection of binary objects, watershed segmentation using layer-stripping for marker identification and applied length approximation to partially-crossed pairs. The algorithm was successfully tested on three populations where binary projections from real single crystals were randomly overlapped to create three populations with either pairs, trios and quintets. The three populations consisted of \( \approx 19000 \) particles to ensure all the overlap cases and configurations were tested. The crystals were prepared using cooling crystallization, but were divided into three groups which had undergone different treatment conditions of temperature cycling and wet-milling to obtain as many size and aspect ratios as possible. The size of the needle-like particles was approximated using an area conserved box that removes surface abnormalities and minimizes size overestimation.

Visual evaluation showed good separation even in extreme overlapped cases and the statistical analysis showed high separation accuracy at, first, low overlap intensity with 91.6% of particles detected. For the detected particles, the length and width were correctly estimated (within 10% error) for 86.7% and 78.6% of the particles, respectively.

As expected, the accuracy decreases with increasing overlap intensity, but remains satisfactory even for the population composed solely of extreme overlaps (quintets), where 71.9% of the particles are detected. For the detected particles, in this case, the length and width were correctly estimated for 74.6% and 77.7% of the particles. In addition, the properties of the particle size and shape distribution, such as the average and standard deviation of length and width were accurately measured from these overlapped cases.

While the presented algorithm is very effective for needle-like particles, it is indeed specialized and limited to separating and sizing such objects and would perform poorly on objects that are not defined by two parallel edges, e.g., round overlapping objects or objects with entirely different characteristics. One avenue to resolve this would be to develop further branches of methodology that could be used to deal with non needle-like morphologies along with a pre-classification step to decide which algorithm to use on the shapes at hand, as other authors have shown previously [25, 26].

The off-line motorized-stage equipment used allows controlling the overlap intensity by decreasing the amount of powder dispersed on the plexiglas plate. Nevertheless, the ability to separate and measure overlapped particles instead of rejecting them, gives the advantage of decreased number of camera frames required to measure a sample (a sample large enough to be representative of a whole population), hence the measurement time decreases. Regardless of the fact that the separating method was tested on off-line equipment, the separation of overlapped crystals is useful for in-process measurement equipment, as well. In fact, in that context, the algorithm is arguably of higher importance, because the suspension density and therefore the particle overlap is not independently tunable in-situ without affecting
the performance of the process (e.g., productivity) or by introducing sampling loops with dilution capabilities.

The proposed algorithm can also be combined with other image analysis methods, which can identify if the separated overlapped needle-like particles are aggregated or not [27, 28]. One could envision that the present algorithm is useful also in the context of sizing primary particles contained in agglomerates/aggregates (i.e., particles that are connected via bridges, not just overlapped on an image), enabling a more in-depth analysis of agglomeration processes than before [32, 33]. Furthermore, the presented algorithm can potentially assist with accurate parameter estimation in mathematical models used to predict [57] and control [58] the evolution of needle-like particle size and shape distributions through processes such as growth, breakage and dissolution.

Acknowledgements

This work was financially supported by EPSRC (Voucher number: 17100002), MSD and AstraZeneca. The authors are thankful to Dr Lorenzo Codan (MSD) and Dr David Wilson (AstraZeneca) for the fruitful discussions and want to acknowledge Dr Giulio Perini (formerly University of Manchester) for his assistance in the lab, as well as, his suggestions for analysis of the results.

References


