Dynamic Data-Driven Distribution Tracking of Nanoparticle Morphology

Chiwoo Park[1(✉)] and Yu Ding[2]

[1] Florida State University, Tallahassee, FL 32310, USA
cpark5@fsu.edu
[2] Texas A&M University, College Station, TX 77840, USA
yuding@tamu.edu

Abstract. We present a data-driven distribution tracking system that is capable of tracking the process quality in a chemical synthesis process for nanoparticles. In the process, the process quality is defined as a distribution of particle sizes and shapes, which influence the functionalities of nanoparticles. A system of tracking the distribution of nanoparticle sizes and shapes consists of three components: (a) *in situ* measurement system, (b) a mathematical model to represent nanoparticle sizes and shapes, their distributions and the temporal changes in the distributions, and (c) a statistical algorithm to estimate the model with *in situ* measurements. We will review the state-of-the-art approaches to tracking the time-varying distribution of particle sizes and shapes. The advance of the distribution tracking by combining complementary *in situ* instruments based on the DDDAS paradigm is discussed.

Keywords: Shape model · Distribution tracking · *in situ* metrology

1 Introduction

Nanoparticles are minuscule particles whose dimensions are less than 100 nm. The functional properties of nanoparticles are heavily influenced by their sizes and shapes, so one can fine-tune the functionalities by simply changing the sizes and shapes. The relation of nanoparticles to sizes and shapes has been studied for many promising applications. For example, the dependency of the surface plasmon property of metal nanoparticles on the particle's sizes was studied for photo-thermal destruction of cancer cells [2], and semiconductor nanoparticles of various sizes were tested as catalysts to promote carbon nanotube growth [9].

A promising method of producing nanoparticles in large quantities is a chemical growth process, known as the self-assembly process [1]. In the chemical growth, atoms and molecules are added to a reaction solution, and those small-scale objects randomly collide in the solution, following a diffusion or a random

We acknowledge support for this work from the AFOSR (FA9550-13-1-0075, FA9550-16-1-0110, FA9550-18-1-0144).

© Springer Nature Switzerland AG 2020
F. Darema et al. (Eds.): DDDAS 2020, LNCS 12312, pp. 132–139, 2020.
https://doi.org/10.1007/978-3-030-61725-7_17

Brownian motion. Some collisions could lead to merging or aggregations. The small-scale objects are aggregated to become larger nanoparticles through multiple stages of mergers. The growth process is influenced by many individualized and localized factors, including the movements of individual objects, local densities of small-scale objects, and the frequency and effectiveness of individual collisions. Because of all these, every single nanoparticle exhibits a unique growth, so that the final sizes and shapes of the nanoparticles resulting from a growth process are unlikely equal but are rather likely form a distribution of a wide span. Producing nanoparticles with a concentrated distribution in both size and shape has been long desired by materials scientists [8].

Producing nanoparticles with controlled sizes and shapes has been attempted experimentally [3], which is to repeat the cycles of trying different chemical recipes and then check particle outcomes by the means of imaging. However, a pure checking of the sizes and shapes of the final nanoparticles does not give any clue on why the outcomes are bad nor guide process improvement. We believe that tracking the sizes and shapes of nanoparticles in the transient period of a growth process provides a crucial clue on how the growth progresses. This paper introduces the problem of tracking the evolution of nanoparticle sizes and shapes, as represented in a time-varying dynamic distribution, and review the state-of-the-art approaches. The tracking problem discussed in this paper is different from the object tracking problem in computer vision [4,5,14], which seeks the trajectories of individual objects and their characteristics instead of the distribution of the characteristics. Our review in Sect. 2 is only focused on the problem of tracking the distribution. The advance of the distribution tracking by combining complementary *in situ* instruments based on the dynamic data driven application system (DDDAS) paradigm is discussed in Sect. 3.

2 Distribution Tracking of Nanoparticles

A system of tracking the time-varying distribution of nanoparticle sizes and shapes consists of three components: (a) *in situ* measurement system, (b) a mathematical model to represent nanoparticle sizes and shapes, their distributions and the temporal changes in the distributions, and (c) a statistical algorithm to estimate, in near real-time, the distribution models with *in situ* measurements. Section 2.1 introduces existing approaches on distribution tracking for both size and shape, and Sect. 2.2 reviews on distribution tracking for size only.

2.1 Shape Distribution Tracking

Measurement Instrument. For tracking nanoparticle shapes in time, an *in situ* imaging of nanoparticles at a nanometer spatial resolution is necessary. There are many microscopic imaging techniques with nano-meter spatial resolutions such as electron microscopes. Most of them had not equipped with *in situ* imaging capability, mainly because wet material samples from chemical processes running in liquid phases cannot be placed on the high vacuum environment of

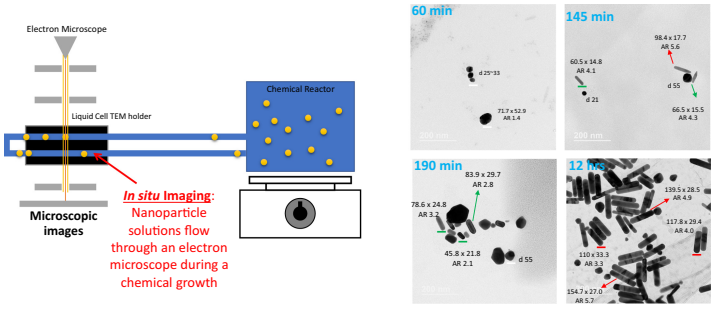
a microscope sample chamber. In order to image those wet samples in the conventional microscopes, special sample holders should be attached for adding *in situ* capability. For examples, a liquid cell sample holder have a very thin layer of liquid samples sandwiched by two silicon or graphene windows [18], and the windows isolate wet samples from the vacuum environment.

As illustrated in Fig. 1-(a), a micro-tubing can be placed to connect in between a reaction chamber and the liquid cell, through which a reaction solution is continuously pumped into the liquid cell. Therefore, the liquid cell will be replenished continuously with reaction solutions taken at different times of a chemical nanoparticle growth process, and taking microscope images of the samples in the liquid cell with an imaging interval would generate a sequence of images containing nanoparticles taken at different stages of the growth process.

Shape Model. Each of the images generated by an *in situ* microscope are analyzed to extract the outlines of nanoparticles in the images, using the state-of-the-art image segmentation approaches [10,12,16,19]. Each of the outlines does not only the size and shape information of the corresponding nanoparticle but also includes the pose of the nanoparticle, where the 'pose' implies the orientation and location of the nanoparticle in the image. The size and shape information can be achieved by discarding the pose information from the outline.

There are quite a few existing works in the shape modeling for nanoparticles. Here we introduce one modern approach. Park [11] represented the outline as a closed curve. A closed curve in $\mathbb{R}^2$ has the circular topology $\mathbb{S}^1$. Therefore, a closed curve can be represented by a parametric curve $\boldsymbol{\phi} : \mathbb{S}^1 \mapsto \mathbb{R}^2$, where the parameter $\theta \in \mathbb{S}^1$ indicates a point on the closed curve, and $\boldsymbol{\phi}(\theta) \in \mathbb{R}^2$ represents the coordinate of the point. To discard the location information from the curve, the closed curve is converted to the centroid distance function, $r : \mathbb{S}^1 \mapsto \mathbb{R}^+$,

$$r(\theta) = ||\boldsymbol{\phi}(\theta) - \boldsymbol{c}_{\phi}||,$$



(a) dynamic electron microscope (b) sequence of dynamic electron microscope images

Fig. 1. *In situ* electron microscope for measuring nanoparticle sizes and shapes at different stages of a nanoparticle growth process. Panel (a) shows a flow-through system attached on a conventional electron microscope that enables a realtime imaging, and panel (b) shows exemplary images from the system.

where $||\cdot||$ is the L2-norm, and $\boldsymbol{c}_\phi = \frac{1}{2\pi}\int_{\mathbb{S}^1}\boldsymbol{\phi}(\theta)d\theta$ represent the centroid of the closed curve. Let $\mathcal{R}$ represent a collection of all such centroid distance functions. The centroid distance function $r \in \mathcal{R}$ still contains the orientation of the corresponding closed curve. Let $\gamma : \mathbb{S}^1 \mapsto \mathbb{S}^1$ denote a diffeomorphism from $\mathbb{S}^1$ to $\mathbb{S}^1$ with a constant first derivative, and let Γ denote the space of all such diffeomorphisms. An $\gamma \in \Gamma$ defines a group action on $r \in \mathcal{R}$ in that $r \circ \gamma$ belongs to $\mathcal{R}$. In fact, the group action rotates the centroid distance function $r \in \mathcal{R}$. Therefore, the shape of a centroid distance function of $r \in \mathcal{R}$ can be represented as all rotational variants of r,

$$[r] = \{r \circ \gamma; \gamma \in \Gamma\}, \tag{1}$$

and the space of shapes can be defined as the quotient space, $\mathcal{R}/\Gamma$. The rotationally invariant distance of two shapes $[r_1]$ and $[r_2]$ in $\mathcal{R}/\Gamma$ is defined as

$$d_{\mathcal{R}/\Gamma}([r_1],[r_2]) = \min_{\gamma\in\Gamma}|(r_1 \circ \gamma) - r_2|,$$

where $|\cdot|$ is the L2-norm in $\mathcal{R}$. A centroid distance function r can be rotationally aligned to a reference centroid distance function $r_* \in \mathcal{R}$ by the partial Procrustes alignment with the distance $d_{\mathcal{R}/\Gamma}$, and the aligned r is achieved as $\tilde{r} = r \circ \tilde{\gamma}$, where $\tilde{\gamma} = \arg\min_{\gamma\in\Gamma}||(r_i \circ \gamma) - r_*||$, and $\tilde{r}$ is used as the shape representation.

The shape of a nanoparticle is represented as a rotation aligned centroid distance function $\tilde{r}(\theta)$, and the shape evolution during a chemical growth process can be represented as a time series $\tilde{r}(\theta, t)$, which represents the shape observed at time t. Park [11] used the spline representation of the time series,

$$\tilde{r}(\theta,t) = \sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{m,n}a_m(t)b_n(\theta), t \geq 0 \text{ and } \theta \in [0, 2\pi),$$

where $a_m(t)$'s and $b_n(\theta)$'s are uniform B-spline basis functions with corresponding random coefficients $\alpha_{m,n}$, and M and N are tuning parameters controlling the number of the spline basis functions used. The vectorial representation of the model is

$$\tilde{r}(\theta,t) = (\boldsymbol{b}_\theta^T \otimes \boldsymbol{a}_t^T)\boldsymbol{\alpha},$$

where $\boldsymbol{\alpha} = (\alpha_{1,1},\ldots,\alpha_{M,1},\ldots,\alpha_{1,N},\ldots,\alpha_{M,N})^T$, $\boldsymbol{a}_t = (a_1(t),\ldots,a_M(t))^T$, and $\boldsymbol{b}_\theta = (b_1(\theta),\ldots,b_N(\theta))^T$. Park [11] pointed out that nanoparticles grow in size, so $\tilde{r}(\theta, t)$ should monotonically increase in time t. Let $\mathcal{Q}$ represent the set of all $\boldsymbol{\alpha}$ values to ensure the monotonicity given the fixed basis matrix $(\boldsymbol{b}_\theta^T \otimes \boldsymbol{a}_t^T)$. The unknown coefficient vector $\boldsymbol{\alpha} \in \mathcal{Q}$ defines the temporal evolution of a nanoparticle. The variation in the temporal evolution among multiple nanoparticles from the same growth process can be modeled by posing a probability distribution on $\boldsymbol{\alpha}$. The truncated multivariate normal distribution can be defined,

$$\boldsymbol{\alpha} \sim \mathcal{N}_{\mathcal{Q}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $\mathcal{Q}$ is a support of $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$ is the mean, and $\boldsymbol{\Sigma}$ is the covariance. More generally, one can use a nonparametric distribution such as a mixture,

$$\boldsymbol{\alpha} \sim \sum_{k=1}^{K} \beta_k \mathcal{N}_{\mathcal{Q}}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),$$

where $\beta_k \geq 0$ is the mixture weight satisfying $\sum_{k=1}^{K} \beta_k = 1$. From the probability model, the probability distribution of $\tilde{r}(\theta, t)$ can be induced as

$$\tilde{r}(\theta, t) \sim \sum_{k=1}^{K} \beta_k \mathcal{N}_{\mathcal{Q}}((\boldsymbol{b}_\theta^T \otimes \boldsymbol{a}_t^T)\boldsymbol{\mu}_k, (\boldsymbol{b}_\theta^T \otimes \boldsymbol{a}_t^T)\boldsymbol{\Sigma}_k(\boldsymbol{b}_\theta \otimes \boldsymbol{a}_t)). \tag{2}$$

For a fixed time t, it represents a probability distribution of nanoparticle sizes and shapes at time t. With the time t varying, it represents the temporal evolution of the probability distribution.

Statistical Algorithm. Suppose that there are N_t nanoparticles observed from the microscope image taken at time $t = 1, \ldots, T$. Let $\tilde{r}_{jt}(\theta)$ represent the rotationally aligned centroid distance function for the outline of the jth nanoparticle observed at time t. All the observations are $\mathcal{D} = \{\tilde{r}_{jt}(\theta); j = 1, \ldots, N_t, t = 1, \ldots, T\}$. Given the data, we want to estimate the distribution parameters $\{(\beta_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k); k = 1, .., K\}$ of the mixture model (2). The expectation maximization algorithm would be a natural choice for the mixture model, if K is known. If K is unknown, a possible solution would be to use a model selection criterion such AIC and BIC to choose K, or a fully Bayesian approach can be taken to consider K as an unknown random variable. Park presented the exact Gibbs sampler for the posterior estimation of K along with the distribution parameters. For more details, please refer to the original paper [11].

2.2 Size Distribution Tracking

Measurement Instrument. When it comes to particle size, scattering techniques are more convenient and practical than microscope techniques. The scattering light techniques come with simpler sample preparation and data analysis steps than microscopic imaging. In addition, scattering machines can be loaded with a much larger volume of nanoparticle solution per each measurement than microscope techniques. Accordingly, the size distribution attained using the scattering techniques can base on a larger sample, so as to better represent the size distribution of the whole reaction solution.

One of the most commonly used scattering techniques for particle sizing is the dynamic light scattering. A sample solution is loaded into a dynamic light scattering machine, and a beam of lights is shot on the sample solution. The light beam is scattered by nanoparticles in the sample solution, and the intensities of the scattered light change in time due to the Brownian motion of nanoparticles in the solution. The autocorrelation of the temporal changes in the intensities

is related to the sizes of the nanoparticles in the solution. The autocorrelation function can be analyzed to reveal the distribution of particle sizes in the form of a histogram. For more details, please refer to a relevant work [7].

Size Model. Let $x \in \mathbb{R}^+$ represents the size of a nanoparticle, and let $p_t(x)$ denote the probability density of the size at time t of a nanoparticle growth process. A simple and practical model for $p_t(x)$ may be a log-normal distribution,

$$p_t(x) = \frac{1}{\sqrt{2\pi\sigma_t^2}x} \exp\left\{-\frac{(\log x - \mu_t)^2}{2\sigma_t^2}\right\},$$

where $\mu_t \geq 0$ and $\sigma_t^2 \geq 0$ are the mean and variance of $\log x$. It has been popularly used for representing particle size distributions [6]. The simple parametric model is not good enough when $p_t(x)$ has multi-modalities, i.e., the density function has multiple local maxima. In that case, a non-parametric distribution such as a histogram can be used. Qian et al. [15] modeled the penalized B-spline model to represent the log probability density, $\log p_t(x) = \sum_{j=1}^{n} \alpha_{jt} B_j(x_i)$, where $B_j(x)$ is the jth B-spline basis function, and α_{jt} is the corresponding B-spline coefficient. The corresponding density of the size distribution is

$$p_t(x) = q_t \exp\left\{\sum_{j=1}^{n} \alpha_{jt} B_j(x_i)\right\}, \tag{3}$$

where $q_t > 0$ is a normalizing constant. The unknown coefficient vector, $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{nt})^T$, parameterizes the particle size distribution at time t, and the temporal change in $\boldsymbol{\alpha}_t$ characterizes the temporal evolution of the particle size distribution. The coefficient vectors can be spatially and temporally correlated. The consideration of the spatial and temporal correlation will be considered by means of incorporating the regularization terms in the statistical algorithms that will be discussed in the next section.

Statistical Algorithm. Suppose that nanoparticles undergoes a nanoparticle growth process, which makes the particle size change following the model (3), and dynamic light scattering measurements are achieved for the samples of nanoparticles taken from the process at time $t = 1, \ldots, T$. The measurement taken at time t can be analyzed by the existing scattering data analysis algorithm [7], and the outcome of the algorithm is a histogram of particle sizes at time t,

$$\boldsymbol{Y}_t = (Y_{1t}, Y_{2t}, \ldots, Y_{mt}),$$

where Y_{it} represents the number of nanoparticles whose sizes range in the ith histogram bin, $[x_i - \delta, x_i + \delta]$. Each of the bin counts is naturally modeled as a Poisson random variable, $Y_{it} \sim Poisson(\lambda_{it})$, where the Poisson intensity $\lambda_{it} = p_t(x_i)$ is proportional to the sampling density p_t. The log likelihood is

$$L(\boldsymbol{\alpha}_t) = \sum_{i=1}^{m} Y_{it} p_t(x_i) - \sum_{i=1}^{m} \exp(p_t(x_i)).$$

Qian et al. [15] proposed to estimate all distribution parameters $\{\boldsymbol{\alpha}_t; t = 1, ..., T\}$ jointly by maximizing the penalized log likelihood,

$$L(\{\boldsymbol{\alpha}_t; t = 1, ..., T\}) = \sum_{t=1}^{T} L(\boldsymbol{\alpha}_t) + \lambda \mathcal{P}(\{\boldsymbol{\alpha}_t; t = 1, ..., T\}), \tag{4}$$

where $\mathcal{P}(\{\boldsymbol{\alpha}_t; t = 1, ..., T\}) = \sum_{t=1}^{T} \sum_{j=1}^{n} \eta(\alpha_{jt} - \alpha_{j(t+1)})^2 + (1 - \eta)(\alpha_{jt} - \alpha_{(j+1)t})^2$ is the smoothness penalty to ensure that the coefficient values do not have sudden jumps , and λ is a positive constant to determine the degree of the smoothness penalty. Qian et al. [15] proposed the alternating directional multiplier method (ADMM) algorithm to optimize the penalized likelihood function.

The penalized likelihood maximization is solved when the scattering measurements from time $t = 1$ to T are available. Therefore, the distribution is estimated after the whole growth process is completed. Qian et al. [17] proposed an online estimation algorithm to estimate $\boldsymbol{\alpha}_t$ incrementally as soon as the measurements up to time t are available instead of waiting until all the measurements are collected. Qian et al. [17] used an autoregressive model to model the time-varying coefficient vector, $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \boldsymbol{\epsilon}_t$, where $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \sigma_t^2 \boldsymbol{I})$. With the autoregressive model, we would have a hidden Markov model linking $\{\boldsymbol{\alpha}_t; t = 1, \ldots, T\}$ with the Poisson observation model (4). The online estimation algorithm of the Kalman filter type can be used to estimate the hidden Markov model [17].

3 Conclusion

This paper discusses the problem of tracking the time-varying distribution of particle sizes and shapes at different stages of a chemical growth process of nanoparticles. If the distribution can be tracked in realtime, it can be exploited for monitoring the growth process, a prerequisite leading to potential control of nanoparticle growth that produces nanoparticles with desirable sizes and shapes. The major challenges in achieving this goal are whether one can take the size and shape measurements in realtime during a growth process, how one effectively models the distributions of sizes and shapes, and how the mathematical model can be estimated as fast as the realtime measurements arrive. We review the recent developments addressing the three challenges. When only the particle sizes are concerned, quick scattering measurements followed by an online density estimation algorithm [17] can carry out a near real-time tracking of particle size distributions. When both shapes and sizes are concerned, realtime online distribution tracking is not yet available. Addressing this latter problem appears much more challenging, due to the high complexities in dealing with shapes. This challenge can be alleviated by a dynamic data-driven application systems (DDDAS) approach making use of multiple measurement instruments of complementary spatio-temporal resolutions. With multi-resolution instruments, one can primarily track the size distribution in realtime using a temporally fast instrument (e.g., the scattering light techniques), while triggering the estimate of shape distribution only when it is necessary [13].

References

1. Grzelczak, M., Vermant, J., Furst, E.M., Liz-Marzán, L.M.: Directed self-assembly of nanoparticles. ACS Nano **4**(7), 3591–3605 (2010)
2. Hirsch, L.R., et al.: Nanoshell-mediated near-infrared thermal therapy of tumors under magnetic resonance guidance. Proc. Nat. Acad. Sci. **100**(23), 13549–13554 (2003)
3. Jana, N.R., Gearheart, L., Murphy, C.J.: Seed-mediated growth approach for shape-controlled synthesis of spheroidal and rod-like gold nanoparticles using a surfactant template. Adv. Mater. **13**(18), 1389–1393 (2001)
4. Jia, B., Pham, K., Blasch, E., Chen, G., Shen, D.: Diffusion-based cooperative space object tracking. Opt. Eng. **58**(4), 041607 (2019)
5. Jia, B., Pham, K.D., Blasch, E., Shen, D., Wang, Z., Chen, G.: Cooperative space object tracking using space-based optical sensors via consensus-based filters. IEEE Trans. Aerosp. Electron. Syst. **52**(4), 1908–1936 (2016)
6. Kiss, L., Söderlund, J., Niklasson, G., Granqvist, C.: New approach to the origin of lognormal size distributions of nanoparticles. Nanotechnology **10**(1), 25 (1999)
7. Li, X., Tran, P.H., Liu, T., Park, C.: Simulation-guided regression approach for estimating the size distribution of nanoparticles with dynamic light scattering data. IISE Trans. **49**(1), 70–83 (2017)
8. Liu, L., Liang, H., Yang, H., Wei, J., Yang, Y.: The size-controlled synthesis of uniform mn2o3 octahedra assembled from nanoparticles and their catalytic properties. Nanotechnology **22**(1), 015603 (2010)
9. Nikolaev, P., et al.: Autonomy in materials research: a case study in carbon nanotube growth. NPJ Comput. Mater. **2**(1), 1–6 (2016)
10. Park, C., Huang, J.Z., Ji, J.X., Ding, Y.: Segmentation, inference and classification of partially overlapping nanoparticles. IEEE Trans. Pattern Anal. Mach. Intell. **35**(3), 669–681 (2013)
11. Park, C.: Estimating multiple pathways of object growth using nonlongitudinal image data. Technometrics **56**(2), 186–199 (2014)
12. Park, C., Ding, Y.: Automating material image analysis for material discovery. MRS Commun. **9**(2), 545–555 (2019)
13. Park, C., Ding, Y.: Dynamic data-driven monitoring of nanoparticle self assembly processes. Handbook of Dynamic Data Driven Applications Systems, 2nd eds. submitted (2020)
14. Park, C., Woehl, T.J., Evans, J.E., Browning, N.D.: Minimum cost multi-way data association for optimizing multitarget tracking of interacting objects. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 611–624 (2014)
15. Qian, Y., Huang, J.Z., Ding, Y.: Identifying multi-stage nanocrystal growth using in situ TEM video data. IISE Trans. **49**(5), 532–543 (2017)
16. Qian, Y., Huang, J.Z., Li, X., Ding, Y.: Robust nanoparticles detection from noisy background by fusing complementary image information. IEEE Trans. Image Process. **25**(12), 5713–5726 (2016)
17. Qian, Y., Huang, J.Z., Park, C., Ding, Y.: Fast dynamic nonparametric distribution tracking in electron microscopic data. Ann. Appl. Stat. **13**, 1537–1563 (2019)
18. Ross, F.M.: Liquid Cell Electron Microscopy. Cambridge University Press, Cambridge (2016)
19. Vo, G.D., Park, C.: Robust regression for image binarization under heavy noise and nonuniform background. Pattern Recogn. **81**, 224–239 (2018)