A NEW DETECTION SCHEME FOR MULTIPLE OBJECT TRACKING IN FLUORESCENCE MICROSCOPY BY JOINT PROBABILISTIC DATA ASSOCIATION FILTERING

Ihor Smal, Wiro Niessen and Erik Meijering

Biomedical Imaging Group Rotterdam
Erasmus MC – University Medical Center Rotterdam
Email: i.smal@erasmusmc.nl

ABSTRACT

Tracking of multiple objects in biological image data is a challenging problem due largely to poor imaging conditions and complicated motion scenarios. Existing tracking algorithms for this purpose often do not provide sufficient robustness and/or are computationally expensive. In this paper we propose a new object detection scheme, based on importance sampling from image intensity distributions, and show how it can be easily incorporated into a probabilistic tracking framework based on Kalman or particle filtering. Experiments on synthetic as well as real fluorescence microscopy image data from different biological studies show that the resulting tracking algorithm yields smaller localization errors at much lower execution times compared to other available methods.

Index Terms— Kalman filtering, Bayesian estimation, multiple object tracking, fluorescence microscopy.

1. INTRODUCTION

Current biological studies using time-lapse fluorescence microscopy imaging require analysis of huge amounts of image data. A large-scale analysis of the dynamics of subcellular objects such as microtubules, vesicles, or proteins (Fig.1) cannot possibly be done without automatic tracking tools that can robustly deal with extremely noisy image data. By using such tools, biologists also eliminate the bias and possibly the systematic errors they introduce during manual tracking due to intuitive selection of relatively small subsets of objects of interest that are either nicely imaged or exhibit typical or expected motion patterns. Thus, automatic tracking methods capable of following as many objects as possible and classifying their dynamics, are of major interest [1].

Deterministic tracking approaches, which subdivide the problem into frame-by-frame detection followed by frame-to-frame linking, produce good results for image data having a relatively high signal-to-noise ratio (SNR). For subresolution objects, fitting the point spread function (PSF) of the microscope to the image data gives the best results, but this has been shown to break down at SNR < 4 [2, 3]. Linking of detected objects can be accomplished by using simple distance or shape criteria, or by Kalman filtering [4], but the results depend critically on the detection. Alternative methods, based on spatiotemporal segmentation, usually work well only for small numbers of well separated objects [5].

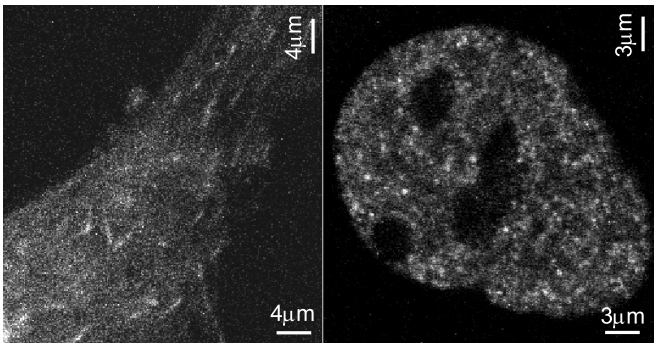

Fig. 1. Examples of subcellular objects to be tracked: microtubules labeled with plus end tracking proteins (left) and proliferating cell nuclear antigens (right), imaged using fluorescence confocal microscopy (the size of the objects is about 200-300nm). The images are single frames from 2D time-lapse studies (courtesy Dr. N. Galjart and Dr. A. Houtsmuller, Erasmus MC).

Recently, probabilistic methods based on sequential Monte Carlo (SMC) approaches, such as particle filtering (PF) [6], have been shown to yield superior robustness and accuracy compared to conventional approaches, but for large numbers of objects they become impractical due to high computational costs. This can be attributed to the large numbers of object state samples (particles) required, each of which needs to be checked against a likelihood. Essentially, this is a "probabilistic detection" scheme, for which more efficient solutions are most welcome.

In this paper we propose a new detection scheme, based on sampling from the image intensity distribution using the so-called h-dome transform from grayscale morphology. This scheme is able to robustly detect objects in highly inhomogeneous backgrounds, which are typical for the applications under consideration (Fig. 1). The same scheme can also be used for track initiation and termination in our tracking algorithm, and gives the possibility to use a bank of Kalman filters, or, to be more precise, to employ the joint probabilistic data association (JPDA) filter as a tracker instead of the PF, which gives enormous computational speed gains.

2. METHOD

Probabilistic tracking is a state estimation problem, where the object state $\mathbf{x}_t$ is estimated in time based on previous states, noisy measurements $\mathbf{z}_t$, and prior knowledge about object properties. Mathematically, it can be formulated as

$$\mathbf{x}_t = f_t(\mathbf{x}_{t-1}, \mathbf{v}_t), \qquad \mathbf{z}_t = h_t(\mathbf{x}_t, \mathbf{u}_t), \tag{1}$$

where f_t and g_t are possibly nonlinear state transition and observation models respectively, and $\mathbf{v}_t$ and $\mathbf{u}_t$ are white noise sources. If the measurement-to-object association is known, (1) can be solved either exactly (when f_t and g_t are linear and $\mathbf{v}_t$ and $\mathbf{u}_t$ are Gaussian) using the Kalman filter, or (in the general case) using SMC approximation methods [7]. The solution is the posterior probability distribution function (pdf) $p(\mathbf{x}_t|\mathbf{z}_1, \ldots, \mathbf{z}_t)$, from which minimum mean square error (MMSE) or maximum a posteriori (MAP) state estimations can be easily computed [7].

Multiple object tracking is complicated by the ambiguous measurement-to-object association problem. The most efficient tracking approaches are the multiple hypothesis tracker (MHT) and the joint probabilistic data association (JPDA) filter [8]. The former builds a tree of hypotheses about all possible measurement-to-track associations, and because of that is not suitable for tracking large numbers of objects. The standard JPDA filter is designed for linear Gaussian models in (1) and uses all measurements to update each track estimate [8]. For practical reasons, measurement gating is often used, which selects for each object the subset of measurements that most likely originated from the object.

Contrary to applications where sensors provide information about the number of objects and their positions, JPDA cannot be applied directly to our applications, because actual position or velocity measurements are not available, but need to be derived from the image data first. Here we extend the JPDA filter for tracking in bioimaging applications and show how appropriate measurements can be created from the image data by using importance sampling techniques [7]. Because of the connection with importance sampling, the proposed filter is termed the IJPDA filter hereafter.

2.1. Detection

We assume that the intensity distribution Z in the images is formed by N_o objects (the intensity distribution of subresolution objects can be modeled by the PSF), background structures (also called clutter) with intensity distribution B, and possibly spatially correlated noise η:

$$Z(\mathbf{r}) = \textstyle\sum_{i=1}^{N_0} I_i \varepsilon(\mathbf{r}; \mathbf{r}_i, \mathbf{\Sigma}_i) + B(\mathbf{r}) + \eta(\mathbf{r}), \tag{2}$$

where $\varepsilon(\mathbf{r}; \mathbf{r}_i, \mathbf{\Sigma}_i) = \exp\left(-\frac{1}{2}(\mathbf{r}-\mathbf{r}_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{r}-\mathbf{r}_i)\right)$, $\mathbf{r} = (x, y, z)^T$, $\mathbf{\Sigma}_i = \text{diag}[\sigma_a^2, \sigma_a^2, \sigma_z^2]$, and $\sigma_a \approx 70$nm, $\sigma_z \approx 250$nm are typical PSF model parameters [3, 6]. The main problem is to accurately estimate the number of real objects N_o and the object positions $\mathbf{r}_i$ in the presence of highly inhomogeneous background structures and noise.

To obtain measurements of object positions and position variances we propose the following procedure. First, the image is 3D Gaussian smoothed at scale σ_m, corresponding to the smallest object of interest. Next, grayscale reconstruction [9] is performed on the smoothed image Z_σ, with mask image $Z_\sigma - h$, where $h > 0$ is a constant. As a result, the image is decomposed into the reconstructed image B_σ and the so-called h-dome image H_σ:

$$Z_\sigma(\mathbf{r}) = H_\sigma(\mathbf{r}) + B_\sigma(\mathbf{r}). \tag{3}$$

The h-dome transform "cuts off" image structures of height h from the top around local intensity maxima, producing "dome"-like structures. In our applications, the h-dome image can be represented as

$$H_\sigma = \textstyle\sum_{i=1}^{N_o} g_{o,i} + \sum_{i=1}^{N_n} g_{n,i} + \sum_{i=1}^{N_b} g_{b,i}, \tag{4}$$

where the functions $g_{o,i}$, $g_{n,i}$, $g_{b,i}$ model the object, noise, and clutter appearance respectively, and

$$g_{o,i} = \begin{cases} 0, \text{ if } (\mathbf{r}-\mathbf{r}_i)^T(\mathbf{\Sigma}+\mathbf{\Sigma}_i)^{-1}(\mathbf{r}-\mathbf{r}_i) > h_c, \\ h - I_{\sigma,i}(1-\varepsilon(\mathbf{r}; \mathbf{r}_i, \mathbf{\Sigma}+\mathbf{\Sigma}_i)), \text{ otherwise.} \end{cases} \tag{5}$$

where $h_c = \log\left(I_{\sigma,i}(I_{\sigma,i}-h)^{-1}\right)$, $\mathbf{\Sigma} = \sigma_\text{m}\mathbf{I}$ with $\mathbf{I}$ the identity matrix, $\max(g_{o,i}) = h$, $a_i = \max(g_{n,i}) < h$ and $b_i = \max(g_{b,i}) \leq h$. Exact analytical expressions for $g_{n,i}$ and $g_{b,i}$ are not available, but $g_{n,i}$ can be approximated by $\varepsilon(\mathbf{r}; \mathbf{r}_i, \sigma_n^2\mathbf{I})$, $\sigma_n \geq \sigma_\text{m}$, and for $g_{b,i}$ a nonparametric representation is assumed.

In our framework, $H_\sigma^s = (Z_\sigma - B_\sigma)^s$ is used as importance sampling function [7], denoted by $q(\mathbf{r}_t|\mathbf{z}_t)$. It describes which areas of the image most likely contain objects. The exponentiation with $s > 1 + \sigma_\text{m}/\sigma_\text{M}$, where σ_M corresponds to the largest object of interest, compensates for the broadening of the original object intensity distributions caused by the Gaussian smoothing, and creates a highly peaked pdf. Because the domains of $g_{o,i}$, $g_{n,i}$, $g_{b,i}$ do not intersect, this exponentiation can be easily computed component-wise in (4). Then, using $q(\mathbf{r}_t|\mathbf{z}_t)$, we draw N position-samples, $\mathbf{r}_t^j \sim q(\mathbf{r}_t|\mathbf{z}_t)$, $j = \{1, \ldots, N\}$, in order to estimate the object positions using Monte Carlo (MC) methods [7]. Next, mean-shifting [10] is used to obtain N_c clusters. For each cluster, the mean position $\bar{\mathbf{r}}_t$ and the covariance $\mathbf{R}_t$ are computed using only the samples belonging to that cluster.

For each cluster, the following two criteria are used to distinguish between a real object and other structures (Fig. 2): (i) the number of samples in the cluster must be larger than the number of samples in the region occupied by the cluster if the sampling would be done from uniform distribution in the whole image domain, and (ii) $\det(\mathbf{R}_t) < (\sigma_\text{m}^2 + \sigma_\text{M}^2)^3/s^3$. These thresholds reflect the fact that for object samples $\mathbf{R}_t \approx (\mathbf{\Sigma}+\mathbf{\Sigma}_i)s^{-1}$ and for background noise samples $\mathbf{R}_t \approx \mathbf{\Sigma}s^{-1}$. Because for the latter the intensity amplitude $a_i \ll h$, the

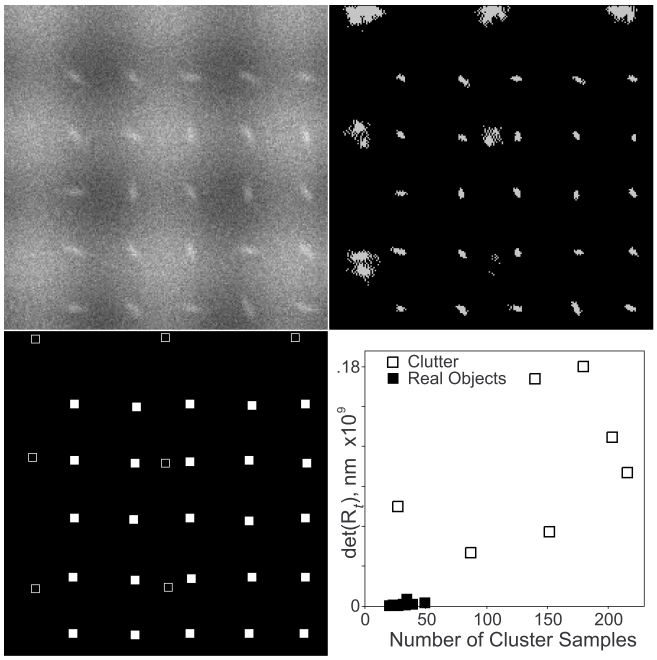

Fig. 2. Synthetic image (top left) showing a number of objects, each having its own orientation and local background structure (brighter regions represent clutter), with noise level SNR ≈ 2, and the results of the described sampling procedure (top right), followed by clustering and classification based on the number of cluster samples and their variances (bottom left and right).

number of samples in the corresponding cluster will be below the mentioned threshold. Clutter, on the other hand, having possibly high intensity values ($b_i \approx h$), produces a large number of samples, but the variance in those clusters is higher than in the case of the largest real object characterized by σ_{M}.

Position estimations of the objects are obtained by computing the center of mass of the cluster samples. An alternative, faster approach is to use only a random subset of the samples to obtain a first estimate up to pixel accuracy, and to refine this estimate by locally fitting a Gaussian model of the PSF to the smoothed image data. A three-point estimation $\hat{x}$ of x based on first estimate x_m is computed as

$$\hat{x} = x_m + (J_{-1,0,0} - J_{1,0,0})/(2J_{-1,0,0} - 4J_{0,0,0} + 2J_{1,0,0}), \quad (6)$$

where $J_{i,j,k} = \log Z_\sigma(x_m + i, y_m + j, z_m + k)$, and similarly for $\hat{y}$ and $\hat{z}$. In the sequel, the implementation that uses this subpixel refinement strategy is denoted as IJPDA$^{\mathrm{sub}}$.

2.2. Tracking

Having the measurements as described above, we can use the JPDA filter to track subcellular objects by assuming a linear Gaussian model for object motion. To estimate the object state $\mathbf{x}_t = (x_t, \dot{x}_t, y_t, \dot{y}_t, z_t, \dot{z}_t)^T$ from the previous estimate $\mathbf{x}_{t-1}$, the Kalman prediction step is used:

$$\mathbf{x}_{t|t-1} = \mathbf{F}\mathbf{x}_{t-1}, \quad \mathbf{P}_{t|t-1} = \mathbf{F}\mathbf{P}_{t-1}\mathbf{F}^T + \mathbf{Q}, \quad (7)$$

with error covariance matrix $\mathbf{P}_t$, process noise covariance matrix $\mathbf{Q} = q_1 \mathrm{diag}[\mathbf{Q}_1, \mathbf{Q}_1, \mathbf{Q}_1]$, and state transition matrix $\mathbf{F} = \mathrm{diag}[\mathbf{F}_1, \mathbf{F}_1, \mathbf{F}_1]$, where

$$\mathbf{F}_1 = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{Q}_1 = \begin{pmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{pmatrix}. \quad (8)$$

Matrix $\mathbf{F}$ models a nearly constant velocity motion with small random accelerations [8]. During the update step, the estimate of $\mathbf{x}_t$ is computed from the predicted state $\mathbf{x}_{t|t-1}$ and the corresponding measurement $\bar{\mathbf{r}}_t$ as

$$\mathbf{x}_t = \mathbf{x}_{t|t-1} + \mathbf{K}_t(\bar{\mathbf{r}}_t - \mathbf{H}\mathbf{x}_{t|t-1}), \quad (9)$$

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{H}\mathbf{P}_{t|t-1}, \quad (10)$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}^T\mathbf{S}_t^{-1}, \quad (11)$$

$$\mathbf{S}_t = \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R}_t N_s^{-1}, \quad (12)$$

and $\mathbf{H} := (h_{ij})_{3\times 6}$ is the observations matrix with all zero elements except $h_{11} = h_{23} = h_{35} = 1$. The measurement $\bar{\mathbf{r}}_t$ for each track is obtained by drawing N_s samples from $q(\mathbf{r}_t|\mathbf{z}_t)$ in the image region defined by the measurement gate $C_t(\mathbf{r}) = \{\mathbf{r} : (\mathbf{r} - \mathbf{r}_{t|t-1})^T\mathbf{S}_t^{-1}(\mathbf{r} - \mathbf{r}_{t|t-1}) < 9\}$, which encompasses the most likely measurements and corresponds to the 3-standard deviation level.

Robust tracking also requires procedures for dealing with track initiation, termination, and interaction. In our new algorithm, initiation of new tracks is accomplished by employing the detection scheme described in Section 2.1, which finds objects at every time step during the tracking and compares them to existing ones. New tracks at time t are initialized with position $\bar{\mathbf{r}}_t$ and a velocity estimated as the difference between the estimate of the position $\bar{\mathbf{r}}_{t+1}$ (computed in the vicinity of $\bar{\mathbf{r}}_t$) and $\bar{\mathbf{r}}_t$. For track termination, the determinant $\det(\mathbf{R}_t)$ for each object is monitored. A track is terminated when, for two successive frames, $\det(\mathbf{R}_t) < (\sigma_{\mathrm{m}}^2 + \sigma_{\mathrm{M}}^2)^3/s^3$, where the threshold corresponds to the situation that the measurement $\bar{\mathbf{r}}_t$ was created from the background structures.

Tracking approaches that assume a one-to-one measurement-to-track assignment (as in most of the deterministic tracking approaches and some of the probabilistic ones), fail to resolve the most ambiguous track interaction scenarios, where two or more objects come close to each other and produce only one measurement for a few time frames. By incorporating prior knowledge about the objects to be tracked (for example, microtubules are rigid structures that cannot easily bend, and because of that their direction of movement before and after the interaction should be approximately the same), we reduce the rate of incorrectly switched tracks that previously caused problems [6].

3. RESULTS

The performance of the proposed tracking algorithm was evaluated using synthetic (with ground truth) but realistic

image sequences containing 20-50 objects resembling microtubules (MTs) and androgen receptors (ARs) moving for 30-50 time frames according to a nearly constant velocity model [8]. Comparison (Table 1) with manual tracking (done by five experts), tracking with commercial software package Volocity (Improvision, Coventry, UK), and our previous PF-based tracking algorithm [6], shows that the proposed IJPDA algorithm is consistently more accurate. For SNR>4, the IJPDA^{sub} implementation yields even more accurate localization estimations. Sample results (visually inspected and confirmed by expert observers) of tracking MTs and ARs in real image sequence are shown in Fig. 3.

SNR	Manual	Volocity	PF	IJPDA	IJPDA^{sub}
2	130 (30)	59	43	35	52
3	114 (28)	42	33	32	45
4	101 (19)	33	27	25	23
5	92 (24)	26	21	18	17
7	70 (5)	22	18	16	14

Table 1. Localization errors (RMSE in nanometers) for manual tracking (with interobserver variabilities in parentheses), Volocity, PF, and the proposed IJPDA and IJPDA^{sub} filters.

As for computational cost, on a regular PC (Intel Core 2 Duo 2.6GHz CPU, 4GB RAM), the current Java implementation of the algorithm requires about 15-20 sec. to process a 2D image sequence ($512 \times 512 \times 50$) with 40-50 moving objects. Overall, the algorithm is 20-100 times faster than previous PF-based algorithms [6]. The execution time is not directly proportional to the number of objects, and the most time consuming tasks (Gaussian smoothing, grayscale reconstruction, and mean shifting) can still be further optimized.

4. DISCUSSION

In this paper we have presented an advanced probabilistic tracking algorithm, which incorporates a new detection scheme that transforms originally nonlinear estimation problems, which can be solved only by employing MC approximations, into a linear problem, for which an efficient solution exists in the form of a JPDA filter. This "simplification" is an improvement, because instead of using approximation techniques for finding the posterior object state distribution, we can now employ a bank of optimal Kalman filters. In addition, the same detection scheme can also be used to deal with track initiation and termination. As the experimental results clearly demonstrate, the new algorithm runs at least an order of magnitude faster than previous PF-based tracking algorithms [6], while yielding similar or even higher accuracy. The proposed scheme can be straightforwardly extended to the interacting multiple model (IMM) filter, which can be used for tracking objects whose behavior is better described by multiple, alternating dynamics models.

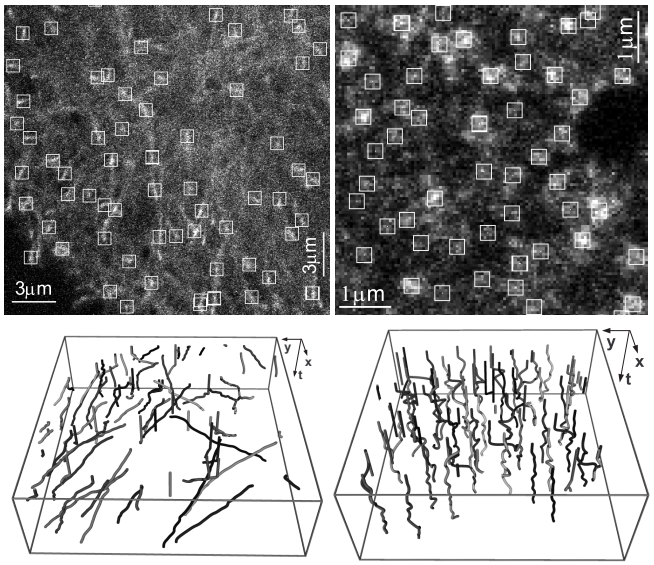

Fig. 3. Single frames from time-lapse fluorescence microscopy images acquired for studying MT (top left) and AR (top right) dynamics, with the objects detected by our algorithm marked by white squares, and the results (bottom left and right) of tracking.

5. REFERENCES

[1] E. Meijering, I. Smal, and G. Danuser, "Tracking in molecular bioimaging," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 46–53, May 2006.

[2] M. K. Cheezum, W. F. Walker, and W. H. Guilford, "Quantitative comparison of algorithms for tracking single fluorescent particles," *Biophys. J.*, vol. 81, no. 4, pp. 2378–2388, Oct. 2001.

[3] D. Thomann, D. R. Rines, P. K. Sorger, and G. Danuser, "Automatic fluorescent tag detection in 3D with super-resolution: Application to the analysis of chromosome movement," *J. Microsc.*, vol. 208, no. 1, pp. 49–64, Oct. 2002.

[4] A. Genovesio, T Liedl, V. Emiliani, W. J. Parak, M. Coppey-Moisan, and J.-C. Olivo-Marin, "Multiple particle tracking in 3-D+t microscopy: Method and application to the tracking of endocytosed quantum dots," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1062–1270, May 2006.

[5] S. Bonneau, M. Dahan, and L. D. Cohen, "Single quantum dot tracking based on perceptual grouping using minimal paths in a spatiotemporal volume," *IEEE Trans. Image Process.*, vol. 14, no. 9, pp. 1384–1395, Sept. 2005.

[6] I. Smal, K. Draegestein, N. Galjart, W. Niessen, and E. Meijering, "Particle filtering for multiple object tracking in dynamic fluorescence microscopy images: Application to microtubule growth analysis," *IEEE Trans. Med. Imaging*, in press.

[7] S. M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[8] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, Norwood, MA, 1999.

[9] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, Apr. 1993.

[10] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.