Available online at www.sciencedirect.com
ScienceDirect
Journal of Structural Biology 159 (2007) 347–358
Journal of Structural Biology
www.elsevier.com/locate/yjsbi

A novel supervised trajectory segmentation algorithm identifies distinct types of human adenovirus motion in host cells

Jo A. Helmuth [a,1], Christoph J. Burckhardt [b,1], Petros Koumoutsakos [a], Urs F. Greber [b], Ivo F. Sbalzarini [a,*]

[a] Institute of Computational Science, ETH Zurich, CH-8092 Zurich, Switzerland
[b] Institute of Zoology, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received 27 October 2006; received in revised form 1 April 2007; accepted 3 April 2007
Available online 14 April 2007

Abstract

Biological trajectories can be characterized by transient patterns that may provide insight into the interactions of the moving object with its immediate environment. The accurate and automated identification of trajectory motifs is important for the understanding of the underlying mechanisms. In this work, we develop a novel trajectory segmentation algorithm based on supervised support vector classification. The algorithm is validated on synthetic data and applied to the identification of trajectory fingerprints of fluorescently tagged human adenovirus particles in live cells. In virus trajectories on the cell surface, periods of confined motion, slow drift, and fast drift are efficiently detected. Additionally, directed motion is found for viruses in the cytoplasm. The algorithm enables the linking of microscopic observations to molecular phenomena that are critical in many biological processes, including infectious pathogen entry and signal transduction.

Keywords: Trajectory segmentation; Trajectory analysis; Motion pattern; Support vector machine; Virus infection; Transport; Computation

1. Introduction

Recent advances in light and video microscopy provide us with unprecedented information about highly complex and plastic biological processes. The acquisition and analysis of trajectories, in particular, is of fast-growing interest and importance in the life sciences. Examples include the characterization of the motion of organelles and signaling molecules in cells (Montero and Heisenberg, 2004; Gonzales-Gaitan, 2003). Intriguing motion patterns have also been revealed by tracking the migration and shape changes in cells of developing and diseased organs (Gasman et al., 2003), or by tracking the invasion of cells and entire organisms by parasites, such as fungi, bacteria, viruses, or neuropathological agents such as prions (for reviews, see Finlay and McFadden, 2006; Marsh and Helenius, 2006; Campana et al., 2005; Chisholm et al., 2006). Viruses infect cells in cohorts, and their trajectories are readily accessible for large-scale analysis (Marsh and Helenius, 2006; Greber and Way, 2006). They trigger multiple independent infectious entry events. These events can be characterized by tracking single fluorescently tagged virus particles (Nakano and Greber, 2000). In the fields of intracellular transport and cell signaling, a number of recent studies have used object tracking. Quantitative description of the trafficking of lipid-bounded organelles, such as endosomes and secretory vesicles, for example, starts to reveal new features of membrane transport in activated and resting cells (Meier and Greber, 2003; Rink et al., 2005).

An important catalyst for motility analysis has been the development of robust, automated particle tracking algorithms with high speed and accuracy, particularly at low signal-to-noise ratios (Gasman et al., 2003; Ponti et al., 2003; Sbalzarini and Koumoutsakos, 2005). In order to

* Corresponding author.
E-mail address: ivos@ethz.ch (I.F. Sbalzarini).
[1] These two authors contributed equally to this work.

1047-8477/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2007.04.003

characterize the resulting trajectories, several global analysis methods on the whole-trajectory level have been proposed. These include Mean Square Displacement (MSD) analysis (Qian et al., 1991), Moment Scaling Spectrum (MSS) analysis (Ferrari et al., 2001; Sbalzarini and Koumoutsakos, 2005; Ewers et al., 2005), as well as approaches based on machine learning techniques such as neural networks (Kinder and Brauer, 1993), self-organizing maps (Owens et al., 2000), and hidden Markov models (Fraile et al., 1998). These global analysis methods have had considerable success. MSS analysis (Ferrari et al., 2001; Sbalzarini and Koumoutsakos, 2005; Ewers et al., 2005) was, for example, used to characterize the motion of plasma-membrane-bound polyoma virus-like particles, revealing that these particles are mostly confined in their motion (Ewers et al., 2005). Analysis of entire trajectories can, however, be misleading when the motion of the object entails different modes. This can be the case for receptor–ligand trajectories, comprising transient phases of directed motion, confinement, and free diffusion (Chen et al., 2004; Kusumi et al., 2005). Spatially resolved analysis methods are required in order to reveal and characterize the mosaic of motion patterns within a single trajectory.

The use of a moving window was proposed as one possible approach to render global analysis methods capable of detecting temporal changes in the mode of motion. While this approach allows resolving changes on the order of the window size, its resolution is limited by the averaging nature of the method, since reducing the window size increases the statistical uncertainty of the values (Qian et al., 1991). In addition, manual thresholding of the moving-window quantities is necessary to detect transitions between motion patterns (Huet et al., 2006).

In this work we present a novel automatic trajectory segmentation algorithm based on supervised support vector classification. We report benchmark tests on synthetic data and demonstrate the suitability of the algorithm for large-scale analysis of the motion patterns of fluorescently labeled human Adenovirus-2 (Ad2) particles in live cells. Adenoviruses cause acute inflammations of the respiratory and ocular mucous membranes and hyperplasia of the submucousal regional lymphoid tissue (Horwitz, 2001), being deadly for immunocompromised persons. The human adenovirus serotype 2 enters cells via receptor-mediated endocytosis and is rapidly released into the cytoplasm (for a review see Meier and Greber, 2003). It requires an intact microtubule cytoskeleton to transport its DNA genome into the nucleus. Initial quantitative analyses of viral motion in the cytoplasm revealed long-range directional transport in normal cells, and short-range transport events in cells treated with the microtubule-depolymerizing drug nocodazole (reviewed in Greber and Way, 2006).

Visual inspection of Ad2 trajectories indicates that they contain different patterns of motion (Fig. 2). We focus on the analysis of patterns that are of non-random nature and may thus relate to certain underlying biological mechanisms. In the case of Ad2 motion, these patterns are: confined motion, slow drift, fast drift, and directed motion. Confined motion is characteristic of receptor-ligand complexes experiencing steric hindrance (Kusumi et al., 2005). Slow drifting motion of filopodial actin filaments is powered by myosin motors and actin dynamics as observed in cultured neuronal cells (Medeiros et al., 2006; Lehmann et al., 2005). Fast drifting and directed motion typically depend on microtubules and motor proteins, which are universal in eucaryotic cells (Vale, 2003; Greber and Way, 2006).

The present algorithm is applied to identifying these motion patterns under different experimental conditions and in different cell types. The algorithm relies on the reduction of the information available by identification of specific features on the virus trajectories. It overcomes many of the limitations of averaging moving-window methods and manual thresholding (Huet et al., 2006), and it enables the detection of short pattern segments and the extraction of physical parameters that describe each pattern. Moreover, we find significant differences in the motion pattern frequencies across different cell types, suggesting that these patterns might serve as specific fingerprints of the virus-cell interactions.

2. Materials and methods

2.1. Cells, viruses, and chemicals

M21 (human melanoma) cells, obtained from Dr. D. Cheresh (Scripps Research Institute, La Jolla, USA), were grown in RPMI 1640 plus 7% clone III serum (Hyclone, PerBio Science, Lausanne, Switzerland) and 1% non-essential amino acids (Sigma, Buchs, Switzerland). Human embryonic retinoblast 911 cells obtained from Dr. A. van der Eb (University of Leiden, The Netherlands) and human embryonic kidney 293 cells (with characteristics of neuronal cells) obtained from ATCC, were grown in DMEM plus 7% clone III serum. Primary human umbilical vein endothelial cells (HUVEC), provided by Dr. Lan Jornot (University of Geneva, Switzerland), were grown in RPMI 1640 plus 15% fetal calf serum and 15 g/ml endothelial growth factor and 90 g/ml heparin (Invitrogen, Basel, Switzerland). For microscopy, cells were grown to 40% confluency on 18 mm glass cover slips (Menzel Gläser, Braunschweig, Germany). Adenovirus serotype 2 was grown, isolated, and subsequently labeled with atto565 (Atto-tec, Siegen, Germany). Nocodazole (Sigma, Buchs, Switzerland) was dissolved in dimethyl sulfoxide (DMSO) and kept in aliquots at −20 °C until use.

2.2. Microscopy

The experiments were conducted with an Olympus IX71 inverted microscope equipped with a triple band pass excitation filter (488 nm/565 nm/647 nm, Chroma), a temperature controlled incubator box (Solent Scientific) and a Beamlok 2060 Ar/Kr mixed gas laser (Spectra Physics). For Total Internal Reflection Fluorescence Microscopy (TIRFM), a 60 × objective was used together with an addi-

tional 1.6× lens, and the laser light was coupled into the microscope condenser (Tillvision) through an optical fiber. Image sequences were recorded during 52 s at an acquisition rate of 50 Hz with a Till-imago-QE CCD camera (TILL Photonics) on a 640 × 480 pixel chip (9.9 × 9.9 μm/pixel) using Tillvision software (TILL Photonics). The total internal reflection angle was adjusted manually to an estimated illumination depth of about 150 nm. For semi-TIRFM experiments, the total internal reflection angle was tuned until cytoplasmic virus particles became visible.

2.3. Trajectory acquisition

We acquired trajectories of both intracellular and plasma-membrane motion of fluorescently labeled Adenovirus serotype 2 particles labeled with atto565 (Atto-tec). The virus particles retained their full infectivity (data not shown, see Nakano and Greber, 2000; Suomalainen et al., 1999). These viruses are herein referred to as Ad2. Image sequences were recorded at a frequency of 50 Hz during 52 s between 30 and 90 min after virus addition on a semi-TIRFM (for intracellular) and between 0 and 60 min for the TIRFM (for plasma-membrane) setup. The videos were processed with a single-particle tracking program (Sbalzarini and Koumoutsakos, 2005), yielding the 2D trajectories of single Ad2. The signal-to-noise ratio of the image sequences was 2.5, yielding a standard deviation of the measured virus positions of 20 nm (Sbalzarini and Koumoutsakos, 2005), in the following referred to as position noise.

2.4. Post-processing procedures

The spatial extent of segments of confined motion (confinement zone size) was estimated based on the point positions within the corresponding trajectory segments. We defined the zone size as the diameter of the circle, centered at the mean position, containing 95% of the point positions in the segment. The speeds of directed motion, fast, and slow drift were defined as the net end-to-end displacement divided by the temporal duration of the segment.

3. An algorithm to detect trajectory segments

3.1. Definitions

A *trajectory* $T = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_x}\}$ is defined by the n_x *positions* $\mathbf{x}_i$ of the tracked object, sampled at discrete time intervals dt. From the positions we can compute the trajectory *steps* $\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$, with $i \in [1,\ n_x - 1]$, leading to an equivalent representation of the trajectory as an ordered sequence of steps:

$$T' = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{n_x-1}\} \quad \text{with } \mathbf{s} \in \mathbb{R}^d. \tag{1}$$

We define a *feature* p as the scalar result of applying a dimensionality *reduction operator* to a sequence of l_w steps, thus:

$$\Psi : \mathbb{R}^{d \times l_w} \to \mathbb{R}, \quad p = \Psi(\{\mathbf{s}_i, \mathbf{s}_{i+1}, \ldots, \mathbf{s}_{i+l_w-1}\}). \tag{2}$$

A vector of multiple features p_i, resulting from multiple operators Ψ_i, is referred to as a *feature vector* $\boldsymbol{f}$. The different types (or modes) c of motion within a trajectory are assumed to be identified by characteristic feature vectors, or, more generally, by a characteristic probability density distribution in feature vector space. If the feature vectors are sensitive to position noise in the trajectories, the point positions can be *filtered* before computing the features: $\widetilde{T} = \Phi(T)$. The *trajectory filter function* Φ hereby describes a unique and explicit mapping from the point positions in T to those in $\widetilde{T}$. Finally, we define a *trajectory part* P as a subset of T' containing all steps between two discrete time points. The process of *trajectory segmentation* recognizes pre-defined motion patterns (defined through their characteristic feature vectors) in the trajectory T, and subdivides T into disjoint trajectory parts P_i, each containing only one type of motion c. Such trajectory parts that only contain a single type of motion are termed *trajectory segments* S.

3.2. Algorithm

The trajectory segmentation algorithm consists of four steps:

(1) Trajectory preprocessing
(2) Feature extraction
(3) Feature classification
(4) Classifier output filtering and label assignment

Step 1: The purpose of *trajectory preprocessing* is to remove position noise, which complicates robust detection of drifting motion (data not shown), from the data by means of a trajectory filter function Φ. For the detection of fast and slow drift, we use the following filter of width w:

$$\Phi : T \to \widetilde{T}, \qquad \tilde{\mathbf{x}}_i = \frac{1}{w} \sum_{j=(i-1)w+1}^{iw} \mathbf{x}_j, \tag{3}$$

where we choose $w = 10$ for fast drift and $w = 50$ for slow drift, reflecting the typical length and speed of these motion types. The detection of directed and confined motion does not require position noise removal.

Step 2: In the *feature extraction* step, we transform the trajectory parts to low-dimensional feature vectors by choosing reduction operators Ψ_i (Eq. (2)). The features are required to characterize the structure of a sequence of steps in order to identify the motion type. The operators are applied to all possible trajectory parts of length l_w. It is important that all features be invariant

with respect to rigid-body rotations and translations of entire trajectory parts. We use the following seven features:

(a) Net displacement
(b) Straightness
(c) Bending
(d) Efficiency
(e) Asymmetry
(f) Point position skewness
(g) Point position kurtosis

The operator Ψ_1 for computing the *net displacement* for a trajectory part P_j of l_w steps is given by:

$$\Psi_1 : P_j \to p_1, \qquad p_1 = ||\mathbf{x}_{j+l_w} - \mathbf{x}_j||. \tag{4}$$

Straightness and *bending* are measures of the average direction change between subsequent steps. *Straightness* is defined as

$$\Psi_2 : P_j \to p_2, \qquad p_2 = \frac{1}{l_w - 1} \sum_{i=j}^{j+l_w-2} \cos\beta_i \tag{5}$$

and the operator for *bending* is

$$\Psi_3 : P_j \to p_3, \qquad p_3 = \frac{1}{l_w - 1} \sum_{i=j}^{j+l_w-2} \sin\beta_i. \tag{6}$$

In both cases, β_i denotes the signed angle change between step $\mathbf{s}_i$ and $\mathbf{s}_{i+1}$ with changes to the left being attributed a positive sign.

Efficiency relates the squared net displacement to the sum of squared step lengths:

$$\Psi_4 : P_j \to p_4, \qquad p_4 = \frac{||\mathbf{x}_{j+l_w} - \mathbf{x}_j||^2}{l_w \sum_{i=j}^{j+l_w-1} \mathbf{s}_i^2}. \tag{7}$$

Asymmetry was previously proposed as a feature for the detection of directed motion (Huet et al., 2006). We use the 2D equivalent, accounting for non-cylindrically symmetric point distributions, thus:

$$\Psi_5 : P_j \to p_5, \qquad p_5 = -\log\left(1 - \frac{(\lambda_1 - \lambda_2)^2}{2(\lambda_1 + \lambda_2)^2}\right). \tag{8}$$

λ_1 and λ_2 are the eigenvalues of R, the 2D radius of gyration tensor of the set of all points $\mathbf{x}_i \in P_j$. Position *skewness* and *kurtosis* measure the asymmetry and peakiness of the distribution of points in P_j, respectively. For their calculation, the position vectors $\mathbf{x}_i \in P_j$ are projected onto the dominant eigenvector $\mathbf{v}$ of R, yielding scalars $x_i = \mathbf{x}_i \cdot \mathbf{v}$. *Skewness* and *kurtosis* are defined as the third and fourth moment of the set of x_i, respectively, thus:

$$\Psi_6 : P_j \to p_6, \qquad p_6 = \frac{\sqrt{l_w + 1} \sum_{i=j}^{j+l_w} (x_i - \langle x_i \rangle)^3}{\left(\sum_{i=j}^{j+l_w} (x_i - \langle x_i \rangle)^2\right)^{3/2}} \tag{9}$$

and

$$\Psi_7 : P_j \to p_7, \qquad p_7 = \frac{(l_w + 1) \sum_{i=j}^{j+l_w} (x_i - \langle x_i \rangle)^4}{\left(\sum_{i=j}^{j+l_w} (x_i - \langle x_i \rangle)^2\right)^2}. \tag{10}$$

For each motion pattern, a certain subset of features is expected to yield optimal classification performance. We determine these optimal feature subsets by exhaustive search as described in Appendix A.1. We include a total of nine features in the search, but only the seven described above occur in the description of any of the considered motion patterns. The globally optimal feature sets for the four pre-defined motion patterns are:

- Directed motion: $\{p_1, p_2, p_3, p_4, p_5, p_7\}$
- Fast drifting motion: $\{p_1, p_2, p_3, p_5, p_6\}$
- Slow drifting motion: $\{p_1, p_2, p_4\}$
- Confined motion: $\{p_1, p_2, p_3, p_5, p_6, p_7\}$

Step 3: *Feature classification* evaluates whether a feature vector corresponds to a certain class c of motion or not, thus constituting a binary classification problem. In the present work, we use *Support Vector Machines* (SVM) as classifiers. SVM are maximum margin linear classifiers (Schölkopf and Smola, 2002), allowing non-linear classification of feature vectors. The SVM assigns each feature vector an output $o_c \in \{0,1\}$. An output value of 1 indicates that the steps in the trajectory part are very likely to belong to class c, while the opposite is true for an output of 0. Prior to its use, the SVM is trained with samples of known class membership. During this process, the SVM simultaneously minimizes the empirical classification error and maximizes the geometrical margin between the classes (classification robustness). Training samples are generated from manually segmented trajectories. These training data are chosen as follows: (1) they have to cover all the different regimes in feature space; (2) samples of critical parts in the trajectories, such as the transition between different patterns, have to be provided. In practice, the training sets are improved by iterations of training, classifica-

tion, and addition of more samples from misclassified segments. We train one SVM for the detection of each motion type as described in Appendix A.2.

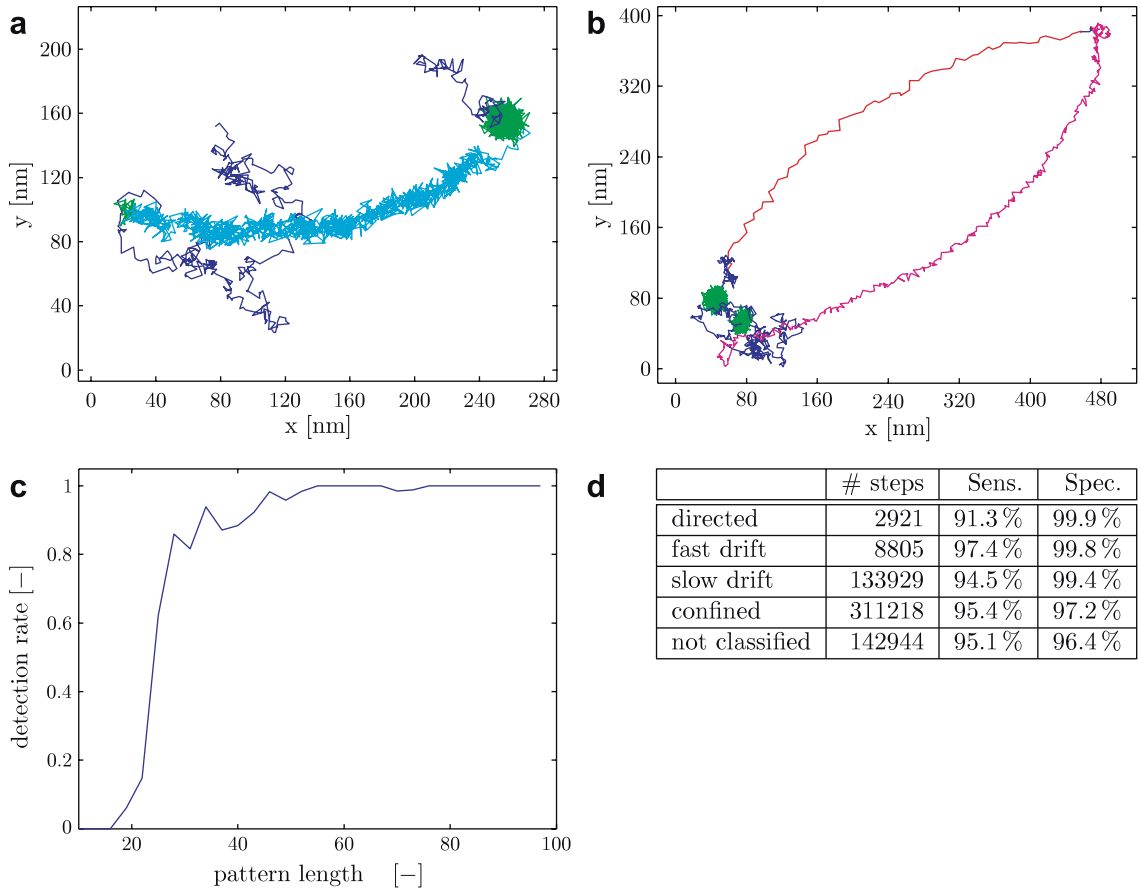
Step 4: The present feature extraction and classification approach raises the following problems: First, due to the overlap of the trajectory parts when computing features, a step can be assigned contradictory classification outputs with respect to a certain pattern. Second, additional conditions such as a minimum length of a trajectory segment may have to be satisfied. Third, the outputs of the classifiers for the different patterns may suggest the affiliation of a step to more than one of the n_c classes, necessitating a decision between classes. These ambiguities are resolved by convolving each sequence of classification outputs o_c with a smoothing kernel, applying a threshold filter to the smoothed sequence, and using class priorities to represent the binary classification tree. This results in a unique class label for each step. Appendix A.3 explains the details of this *classifier output filtering and class label assignment* procedure.

4. Validation of trajectory segmentation on synthetic data

We benchmark and validate the trajectory segmentation algorithm on synthetic trajectories. For each motion type, we measure the rate of correctly classified steps. We demonstrate in addition that the present algorithm is able to accurately detect short trajectory segments, thus overcoming the main limitation of averaging moving-window methods.

4.1. Generation of synthetic trajectories

We apply two different methods to generate synthetic trajectories of particle motion. *Random walk* is used as the standard model for normal single-particle diffusion (Einstein, 1905). Sub-diffusive motion (Saxton, 2001; Weiss et al., 2003) is generated by sampling a sequence of displacements from the *Weierstrass-Mandelbrot function* (Weiss et al., 2003; Berry and Lewis, 1980), independently for the two spatial dimensions. Both methods yield trajectories that are homogeneous with respect to the type of motion. While these trajectories are not suitable for the evaluation of segmentation rates, they are used to test if the motion types occur



	# steps	Sens.	Spec.
directed	2921	91.3 %	99.9 %
fast drift	8805	97.4 %	99.8 %
slow drift	133929	94.5 %	99.4 %
confined	311218	95.4 %	97.2 %
not classified	142944	95.1 %	96.4 %

Fig. 1. (a,b) Application of the trajectory segmentation algorithm to heterogeneous synthetic trajectories. The segments are color-coded as: directed (red), fast drift (magenta), slow drift (cyan), and confined motion (green). Blue trajectory parts represent non-classified motion. (c) Successful detection of motion types is a function of the segment length, as demonstrated here for directed motion. (d) Detection rates for the different motion types in synthetic trajectories (Sens.: sensitivity, Spec.: specificity). The fraction of correctly classified trajectory steps was always higher than 90%.

more frequently in real virus trajectories than in random ones. In order to assess the accuracy of the segmentation algorithm, we create heterogeneous synthetic trajectories that—besides random motion—contain stretches of modeled motion of all four types. The details on the creation of these synthetic data are given in Appendix B.

4.2. Segmentation of synthetic trajectories

Two representative results from the segmentation of heterogeneous synthetic trajectories are shown in Figs. 1a and b. The automated segmentation is in good agreement with manual segmentation. In order to quantify the quality of the segmentation we measure the sensitivity and the specificity for each motion class. This is done on a large set ($N = 200$) of long (4000 steps) synthetic heterogeneous trajectories. The rates are given in the table in Fig. 1d. In all cases, the rates of correct classifications (sensitivity) were >90% with rates of correct rejections (specificity) of >95%. The largest fraction of spurious positives (13%) is observed for fast drifting motion, which has, however, a low frequency (1.5%). Errors mostly occur at transition points between different types of motion. While such transitions were always correctly *detected*, they were not always precisely *localized*. In order to assess the capability of the method to detect short segments, we analyze synthetic heterogeneous trajectories of random walks with intermediate segments of directed motion of varying length. A segment of directed motion is considered detected if at least 80% of the steps are assigned the correct class label. The result is shown in Fig. 1c. Above a length of about 25 steps, the detection rate quickly approaches 100%.

Finally, before applying the algorithm to real Ad2 trajectories, we use homogeneous synthetic trajectories (cf. Section 4.1) to estimate the probability of the four motion types to occur in uniform random walks. This provides confidence intervals for assessing the statistical significance of the Ad2 results. The homogeneous trajectories, representing random walks or sub-diffusive motion, and real Ad2 virus trajectories on the plasma-membrane of M21 cells had the same distribution of step lengths, the same 2-dimensional diffusion constants, and the same MSS slopes (see Supplementary Material for details). The segmentation results are summarized in Table 1. In synthetic random walk trajectories, 98.9% of the steps were correctly classified as random walk, and only minor amounts of the four motion patterns were found. For sub-diffusive synthetic trajectories, the segmentation algorithm yielded a comparable fraction of confined motion (67.2%) as in the Ad2 trajectories (55.9%–65.5%, Table 1), while all other motion types were extremely rare (less than 0.1%).

5. Segmentation of Adenovirus-2 trajectories

We apply the segmentation algorithm to trajectories of Ad2 on live cells. The results shown in Fig. 2 illustrate that the algorithm discriminates close adjacent segments of confined motion (Fig. 2b) and identifies slow drifting motion (Fig. 2c), fast drifting motion (Fig. 2b), and directed motion (Figs. 2a and d). We acquired and analyzed a large number of intracellular trajectories of Ad2 using semi-TIRF recording conditions, either in the absence ($N = 4388$) or presence ($N = 2399$) of the microtubule inhibitor nocodazole (see Section 2). The frequencies of both fast drift and directed motion were strongly reduced in the presence of nocodazole, namely from 2.07% to 0.17%, and from 0.58% to 0.06%, respectively (Table 1). Fast drifts typically lasted for 2–10 s with single-displacement speeds of 20–400 nm/s (Figs. 3e and f). Directed motion was faster (400–1000 nm/s), but lasted for only short periods of 0.3–2.7 s (Figs. 3g and h). The end-to-end displacement of directed motion was nonetheless higher than the one of fast drifting motion. 49% of the fast

Table 1
Frequency analysis of motion types in real Ad2 trajectories compared to synthetic data

	N	Directed		Fast drift		Slow drift		Confined		Not classified	
		t (%)	#	t (%)	#	t (%)	#	t (%)	#	t (%)	#
(*A*) *Synthetic*											
Random walk	3130	0.006 ± 0.003	6	0.36 ± 0.04	102	0.44 ± 0.09	25	0.33 ± 0.02	223	98.9 ± 0.1	3486
Sub-diffusion	3130	0.02 ± 0.01	16	0.04 ± 0.01	11	0	0	67.2 ± 1.0	6155	32.8 ± 1.0	9312
(*B*) *Intracellular*											
No drug	4388	0.58 ± 0.06	209	2.07 ± 0.21	186	2.03 ± 0.38	39	55.9 ± 0.6	5549	39.4 ± 0.6	10305
50 μM Nocodazole	2399	0.06 ± 0.02	19	0.17 ± 0.06	12	1.07 ± 0.40	10	65.4 ± 0.9	3048	33.3 ± 0.9	5476
(*C*) *Cell surface*											
M21	414	0.10 ± 0.04	8	1.29 ± 0.31	26	9.69 ± 1.84	30	57.2 ± 2.0	821	31.7 ± 2.0	1270
293	952	0.003 ± 0.003	1	1.56 ± 0.24	92	14.9 ± 1.2	148	62.1 ± 1.4	2043	21.4 ± 0.8	3113
911	403	0.12 ± 0.06	8	2.05 ± 1.04	22	8.45 ± 1.87	21	57.1 ± 2.4	617	32.3 ± 2.0	1044
HUVEC	361	0.02 ± 0.02	1	1.05 ± 0.40	13	6.62 ± 2.04	14	65.5 ± 2.2	595	26.9 ± 2.0	965

(A) Synthetic Brownian motion is modeled by random walks, sub-diffusive trajectories are modeled using the Weierstrass-Mandelbrot function as outlined in Section 4.1. (B) Motion types in intracellular Ad2 trajectories of human M21 melanoma cells with and without nocodazole. (C) Cell surface motion types of Ad2 on human M21 melanoma, 293 epithelial, 911 epithelial, and primary HUVEC endothelial cells. Bootstrap estimates of standard deviations are indicated by ±.

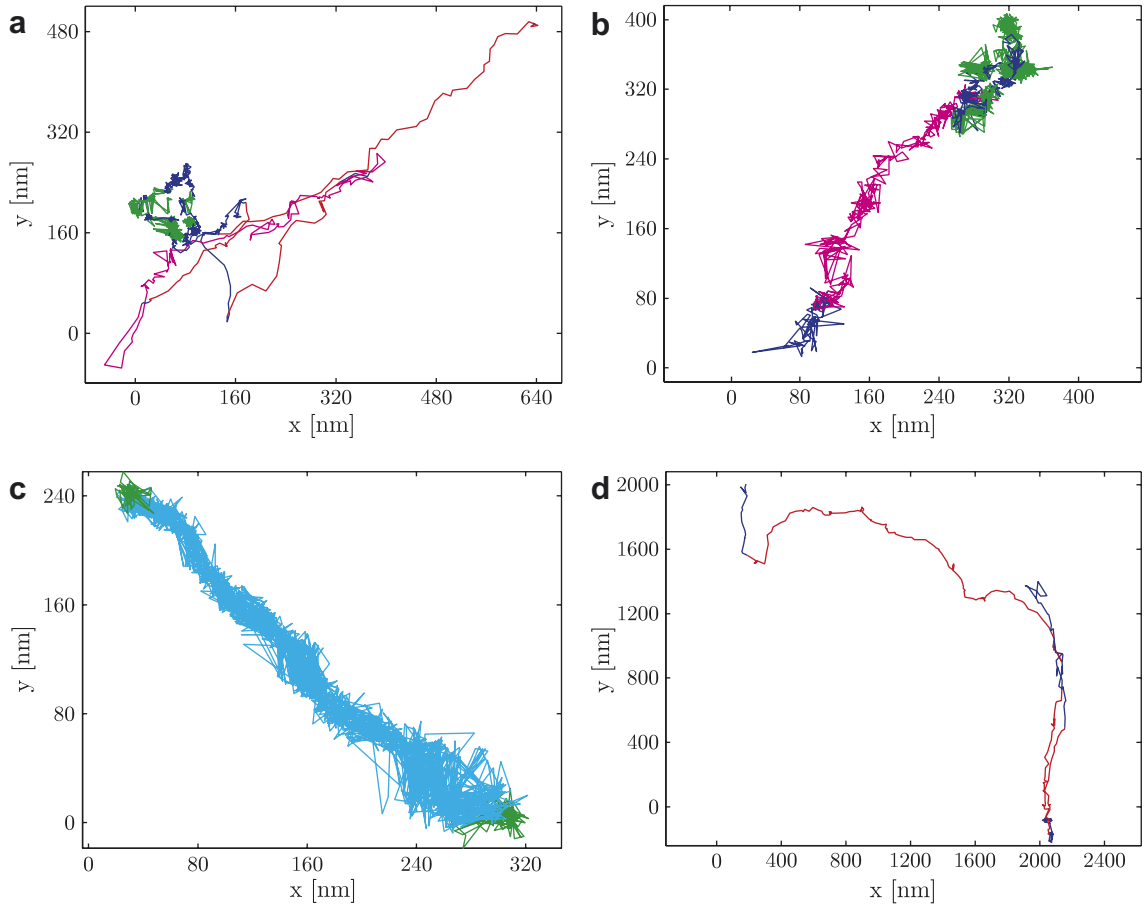


Fig. 2. Segmented intracellular tracks of Ad2 in M21 cells. Segments are color coded: confined motion is green, slow drifting motion is cyan, fast drifting motion is magenta, and directed motion is red. Parts of the trajectories that were not classified are shown in blue. (a, b) Examples of strongly heterogeneous Ad2 trajectories composed of directed motion (not present in b), fast drifting motion, confined motion, and non-classified motion. (c) A segment of slow drifting motion extends over almost the entire trajectory. (d) Example of a trajectory with several separate phases of directed motion.

drifting motion and 30% of the directed motion were present in tracks with overall MSS slopes below 0.5, suggesting a globally sub-diffusive process. This highlights the importance of trajectory segmentation for functional analysis. In addition, fast drift and directed motion were often saltatory and bidirectional (see Fig. 2d), indicative of microtubule-based transport. Besides drastically reducing the frequency of fast drift and directed motion, nocodazole increased the fraction of confined motion from 55.9% to 65.4%, while the amount of slow drift was reduced by 50%. We infer that fast drift and directed motion of intracellular Ad2 depend on intact microtubules, possibly involving motor proteins such as dynein-dynactin or kinesins (Suomalainen et al., 1999; Vale, 2003), or microtubule dynamics (Desai and Mitchison, 1997; Dogterom et al., 2005).

Confined motion was typically short-lived with durations on the order of a few seconds within areas of on average 80 nm diameter (max. 150 nm) (Figs. 3a and b). This high degree of confinement is consistent with earlier observations (Suomalainen et al., 1999). It is possibly attributed to the viscous cytoplasmic environment and the dense cytoskeleton meshwork restricting cytoplasmic diffusion. Alternatively, it could be due to tethering of particles to immobile cellular scaffolds. The slow drifting motion proceeded without interruptions with speeds of 2–15 nm/s (Fig. 3d). These speeds were determined in time intervals of several seconds, comprising hundreds of displacement steps (Fig. 3c). Slow drifting motion often lasted throughout the entire observation period, indicating a high persistence (cf. Fig. 2c).

To further validate the trajectory segmentation algorithm, we used TIRF microscopy to analyze the virus motion on the surface of different cell types within the first 15 min after viral inoculation, including M21 human melanoma cells ($N = 414$), human embryonic kidney 293 cells ($N = 952$), human embryonic retinoblast 911 cells ($N = 403$), and primary human umbilical vein endothelial cells (HUVEC) ($N = 361$). Ad2 particles readily gained access to the basal side of the cells and were amenable to examination by TIRF microscopy. As expected, there was a low fraction of at most 0.12% of directed motion in each cell type (Table 1). The fraction of slow drift varied between 6.6% and 14.9%, while the fraction of confined motion was between 57.1% and 65.5%. Not classified motion accounted for 21.4% to 32.3%. Remark-

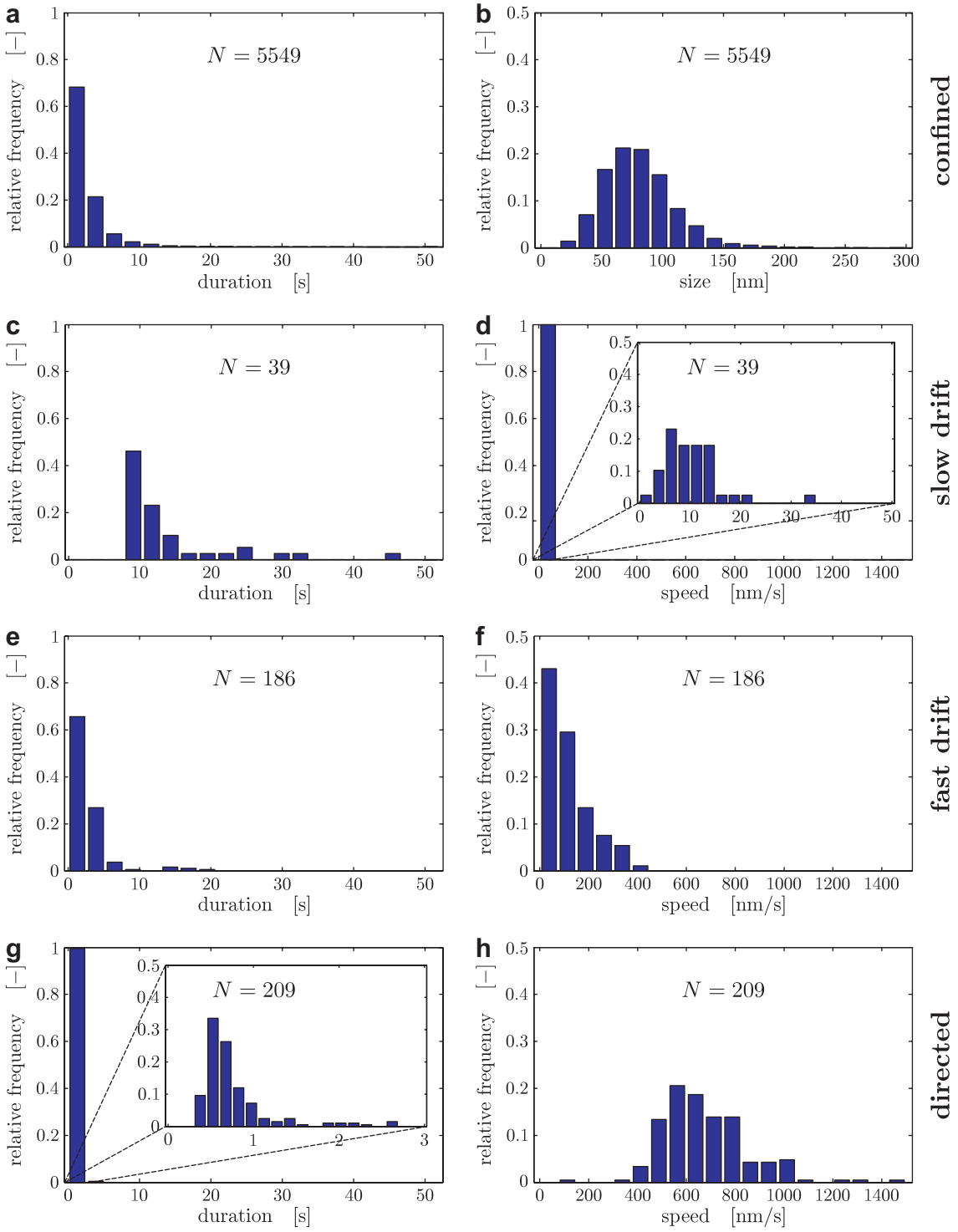


Fig. 3. Analysis of the parameters characterizing intracellular Ad2 motion in M21 cells. Histograms of the duration of segments and their size or speed are represented in panels a, c, e, g, and b, d, f, h, respectively, for confined motion (panels a, b), slow drifting motion (c, d), fast drifting motion (e, f), and directed motion (g, h).

ably, the fraction of slow drift was 3- to 7-fold larger than in trajectories of intracellular virus particles. The fractions of motion types and their standard deviations were estimated by bootstrapping (Hastie et al., 2003). Based on these estimates, the significance of the observed differences in the experimental conditions can be assessed. In the intracellular M21 data the standard deviations of the fractions of fast drift and directed motion were 0.21% and 0.06%, respectively, indicating that the nocodazole-mediated reduction of the observed patterns is statistically highly significant. Significant differences were also observed for cell surface trajectories, most

prominently in the fraction of slow drift and confined motion. The differences in directed motion were negligible due to the small total amount of this motion type. Together, these analyses suggest that the pattern frequencies of the different cell types constitute specific fingerprints. The differences in these fingerprints were significant, since the pattern fractions of the homogeneous synthetic data differed from the experimental data by more than 10 standard deviations. We hence conclude that the motion types that we described in both cell surface and intracellular Ad2 trajectories can not be explained by assuming that a single random process underlies all trajectories. Instead, the data suggest a connection between trajectory segments and Ad2 interactions. The underlying cellular machinery possibly depends on the organization of the plasma-membrane and the cytoskeleton.

6. Discussion

We have presented a novel algorithm that automatically segments different types of motion within trajectories by using support vector classification to identify sets of specific features. Representation of the trajectory parts by feature vectors is essential to overcome the poor classification performance observed on raw point positions (data not shown). We distinguished between: confined motion, slow drift, fast drift, and directed motion. Parts of the trajectories that cannot be associated with any of the motion types are referred to as non-classified motion. Using this approach allowed segmenting synthetic heterogeneous trajectories with only 2–13% of spurious associations of a trajectory step with a motion type, while the sensitivity of the method was above 90% in all cases. We further demonstrated the applicability of the method to the segmentation of trajectories of intracellular virus particles and viruses moving on the surface of live cells. The virus trajectories were reconstructed from high-resolution fluorescence microscope image sequences of live cells, taken at a frame rate of 50 Hz with a signal-to-noise ratio of 2.5. This corresponds to a tracking precision of 20 nm (Sbalzarini and Koumoutsakos, 2005). The presented pattern-based approach enabled us to reliably segment such noisy virus trajectories, where the duration of the different motion types could be as short as 25 steps. This allowed extracting information from trajectories at a time scale that is inaccessible to methods based on averaging moving window statistics.

Previous global statistical analysis of whole virus trajectories suggested that their motion can be, on average, subdiffusive (Ewers et al., 2005). The average motion type was assessed by computing the slope of the MSS, a method that is preferable over MSD analysis due to its higher robustness and accuracy (Sbalzarini and Koumoutsakos, 2005; Ewers et al., 2005; Ferrari et al., 2001). An MSS slope of 0.5 is characteristic for Brownian motion, while values below or above 0.5 indicate sub- or super-diffusion, respectively. The MSS slopes of intracellular Ad2 were measured between −0.14 and 0.86 (although theoretically impossible, negative values occurred due to imaging noise), with a mean of 0.16. Two-dimensional diffusion constants of cytosolic Ad2 have been measured to be in the range of 5.2×10^{-5}–4.6×10^{-1} $\mu m^2/s$ with a mean of 2.6×10^{-3} $\mu m^2/s$ (see Supplementary Material). This is indicative of extensive restriction due to the high viscosity of the cytoplasm (Luby-Phelps, 2000; Sodeik, 2000).

As expected, trajectories that exhibit a large amount of directed motion yielded higher MSS slope values, while lower MSS slopes resulted from high amounts of confinement (see Supplementary Material). Nevertheless, global trajectory analysis failed to identify local variations of motion such as, e.g., short segments of directed motion in otherwise sub-diffusive trajectories. This was demonstrated by the high frequency of directed motion and fast drifts in trajectories with over-all MSS slopes below 0.5. By analyzing the segmentation of synthetic trajectories we were able to show that Ad2 trajectories are highly heterogeneous. They are composed of segments of motion resulting from different biophysical processes, and significant differences exist between cell types.

In conclusion, we have demonstrated that the application of the present algorithm provides insight into the biophysical cause of observed changes in global trajectory parameters under different experimental conditions. Further, important physical parameters such as duration, speed, and length distribution of the different virus motion types could be extracted from the trajectory segments. Information specific to virus trajectories was used to define the features of trajectory segments and to design the trajectory filter function. We anticipate, however, that by adjusting these definitions, the present algorithm can be successfully applied to the segmentation of other types of trajectories, such as from, e.g., movements of whole cells (ORourke et al., 1992; Soo and Theriot, 2005) or organisms (Wolf and Wehner, 2000; Croxall and Silk, 2005), opening a new possibility in the analysis of biological trajectories.

Appendix A. Trajectory segmentation algorithm

A.1. Feature selection

We seek the minimal set of features that is suited to discriminate between the two classes of a given set of training samples. A small number of features is desirable for two reasons: it decreases the computational cost of training and classification, and it yields a more robust classifier by reducing the complexity of the decision boundary. Support vector machines are trained and tested using all possible combinations of the features introduced in Section 3.2. Sensitivity (the proportion of correct classifications of all positive samples) and specificity (the proportion of correct rejections of all negative samples) are used to measure the generalization capability. They are estimated by five-fold cross-validation on the training set. For each feature set

size, the set with the best performance is selected. As we are interested in small sets, the set for which no larger one yielded a significantly improved performance is chosen. This procedure is carried out independently for each pattern.

A.2. Feature classification

We use a C-SVM with radial basis function kernel for feature classification. The parameter values C (penalization of misclassification on the training data) and σ (radius of the kernel function) are determined by grid search for each SVM, i.e., for each motion pattern. The values are given in Table A.1. Classification performance is evaluated by five-fold cross-validation on the training data. The estimated classification performance (as defined in Appendix A.1) is also given in Table A.1.

A.3. Feature extraction, classifier output filtering, and label assignment

Feature vectors are extracted from overlapping trajectory parts of length l_w. Small values of l_w favor detection of short segments and enable more accurate segmentation

Table A.1
Summary of the SVM parameters used, the achieved classification performance on Ad2 training data, and the parameters used in the classification output filters

	SVM parameters							Filter parameters		
	C	σ	N_{pos}	N_{neg}	Sens.	Spec.	Avg.	l_w	t_c	l_{min}
Directed	35.94	0.774	274	1370	0.996	0.994	0.995	25	0.15	20
Fast drift	12.92	3.594	338	1690	0.897	0.985	0.941	10	0.5	8
Slow drift	100	0.464	375	1082	0.859	0.939	0.899	10	0.3	8
Confined	4.46	0.464	2031	2039	0.975	0.956	0.966	50	0.3	30

N_{pos} and N_{neg}: number of positive and negative training samples, respectively; Sens.: sensitivity, Spec.: specificity, Avg.: (Sens + Spec)/2.

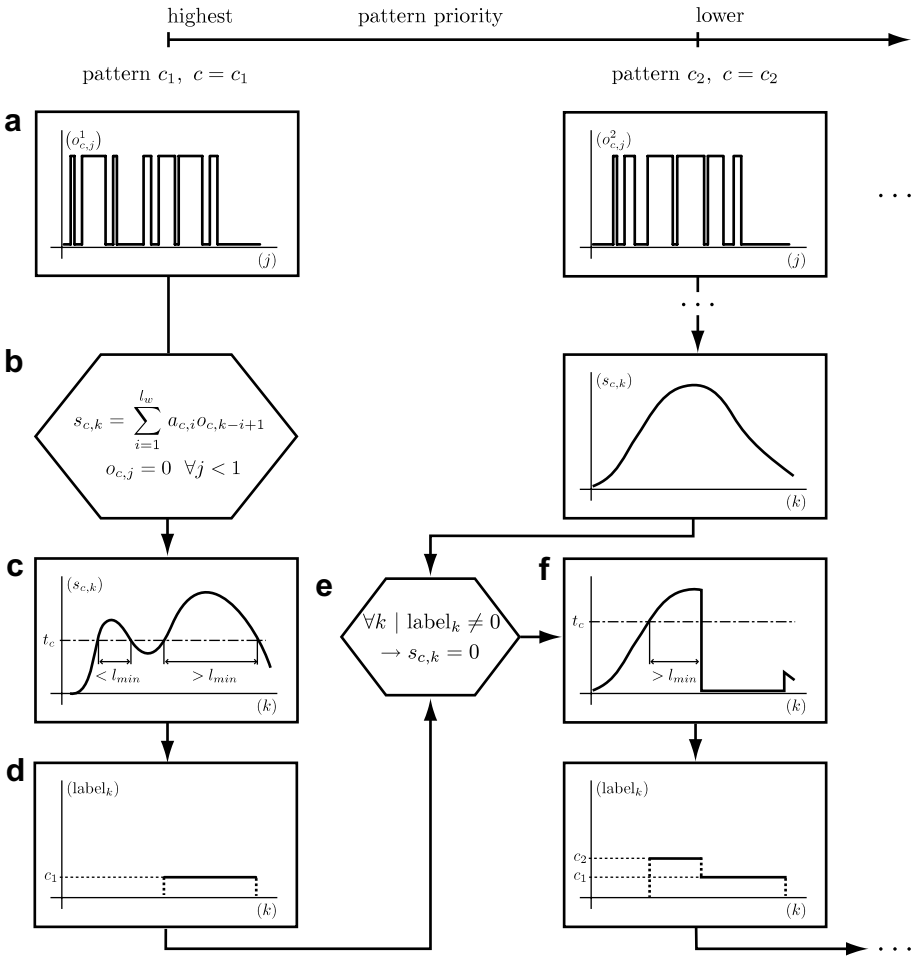


Fig. A.1. Classifier output filtering and class label assignment algorithm. The individual stages for all patterns are chained according to pattern priority (see text for details).

of pattern transitions. Larger values of l_w increase the statistical significance of the detected patterns, i.e., reduce the probability that the same feature vector could be a result of random motion.

The feature vectors are then classified by the trained SVM (cf. Appendix A.2), yielding a series O_c of classifier outputs $o_{c,j} \in \{0,1\}$ for each pre-defined pattern c (Fig. A.1a). Each output is interpreted as a vote for all steps within the corresponding trajectory part. Based on the multiple votes from all class-specific SVM, each step k in the trajectory is assigned scores $s_{c,k}$ that represent the likelihood that this step belongs to pattern c. To this end, we compute a weighted sum of all outputs that are relevant for a particular step (Fig. A.1b), thus:

$$s_{c,k} = \sum_{i=1}^{l_w} a_{c,i} o_{c,k-i+1} \quad k \in [1, n_x - 1] \qquad o_{c,j} = 0 \quad \forall j < 1, \tag{A.1}$$

which amounts to a discrete convolution of the output sequence O_c with a filter kernel A_c:

$$\{s_k\}_c = O_c * A_c. \tag{A.2}$$

In the present work we use a uniform kernel, corresponding to setting all weights $a_i = 1/l_w$, thus equally distributing the outputs among all steps within the trajectory part. Spurious classifications of trajectory parts might, however, add a random component to the sequence of outputs O_c. The problem of assigning the correct label is thus equivalent to finding regions in the score sequence that are higher than the average score level caused by spurious classifications.

A straightforward solution is to use a threshold filter, thus generating a binary signal (either 0 or 1) (Fig. A.1c). The choice of the threshold t_c is hereby constrained by the requirements of the application. A low value is beneficial for precise localization of transitions between patterns. The lowest possible value is given by the expected rate of false positive classifications (e.g. 0.06 for slow drift). A high threshold value improves the separation of short segments of the same pattern and reduces false detections. The highest possible value is given by the expected rate of true positive classifications (e.g. 0.86 for slow drift). The values used in the present Ad2 study are given in Table A.1. We require that the binary signal equals 1 for at least l_{min} steps, in order to reject short and statistically insignificant segments. The binary signal is then converted to a sequence of step class labels by multiplication with c (Fig. A.1d). This can, however, lead to multiple contradictory labels being assigned to the same step.

Such ambiguities are resolved by defining priorities among patterns and processing the classification outputs in order of descending priority. After each pattern, the scores of all steps that have been associated with a pattern are set to zero (Fig. A.1e). In the present work we choose the priorities according to the rank of the average speed of the motion, thus: directed motion before fast and slow drift, before confined motion. All steps that retain a final label of zero are considered not classified. If the trajectory was filtered prior to feature extraction, the label sequence has to be mapped back onto the steps of the original trajectory (not shown in Fig. A.1).

Appendix B. Generation of synthetic data

Homogenous synthetic trajectories are generated using either a simulated random walk or the Weierstrass-Mandelbrot function (Berry and Lewis, 1980; Saxton, 2001; Weiss et al., 2003). In each step of a random walk, the particle is displaced by a value sampled from a two-dimensional Gaussian with mean $\boldsymbol{\mu} = (0,0)^T$ and covariance matrix

$$\mathbf{C} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \qquad \text{with } \sigma^2 = \frac{2D}{f}. \tag{B.1}$$

D is the molecular diffusion constant and f denotes the sampling frequency (image acquisition frame rate). In order to model sub-diffusion, the Weierstrass-Mandelbrot function

$$W_i(t) = A \sum_{n=n_{min}}^{n_{max}} \frac{1}{\gamma^{nH}} (\cos U_i(n) - \cos[\gamma^n t^* + U_i(n)]) \tag{B.2}$$

is sampled independently for all dimensions i at time steps t. $t^* = 2\pi t/t_{max}$, with t_{max} the total time of the synthetic trajectory. Following Saxton (Saxton, 2001), we set $\gamma = \sqrt{\pi}$, $n_{min} = -8$, and $n_{max} = 48$. $U_i(n)$ are uniformly distributed random numbers between 0 and 2π. Computing the cumulative sum of $W_i(t)$ for each dimension yields the point sequence of the trajectory. Any diffusion type (characterized by the diffusion constant D and the MSS slope) can be modeled by adjusting the parameters H and A to match the desired behavior. Confined motion is simulated by adjusting H such that the MSS slope is 0.1, resulting in strong visual similarity to the noisy trajectory of an immobile particle. Directed motion, fast drift, and slow drift are created by superimposing bent directed motions of different speeds (300–600, 20–200, and 8–16 nm/s) and durations (0.5–4, 2–15, and 10–50 s) onto trajectories of confined motion.

Heterogeneous synthetic trajectories are created by assembling stretches of homogeneous trajectories of different motion types.

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2007.04.003.

References

Montero, J.A., Heisenberg, C.P., 2004. Gastrulation dynamics: cells move into focus. Trends Cell Biol. 14 (11), 620–627.

Gonzales-Gaitan, M., 2003. Signal dispersal and transduction through the endocytic pathway. Nat. Rev. Mol. Cell Bio. 4 (3), 213–224.

Gasman, S., Kalaidzidis, Y., Zerial, M., 2003. RhoD regulates endosome dynamics through Diaphanous-related Formin and Src tyrosine kinase. Nat. Cell Biol. 5 (3), 195–204.

Finlay, B.B., McFadden, G., 2006. Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. Cell 124 (4), 767–782.

Marsh, M., Helenius, A., 2006. Virus entry: open sesame. Cell 124 (4), 729–740.

Campana, V., Sarnataro, D., Zurzolo, C., 2005. The highways and byways of prion protein trafficking. Trends Cell Biol. 15 (2), 102–111.

Chisholm, S.T., Coaker, G., Day, B., Staskawicz, B.J., 2006. Host-microbe interactions: Shaping the evolution of the plant immune response. Cell 124 (4), 803–814.

Greber, U.F., Way, M., 2006. A superhighway to virus infection. Cell 124 (4), 741–754.

Nakano, M.Y., Greber, U.F., 2000. Quantitative microscopy of fluorescent adenovirus entry. J. Struct. Biol. 129 (1), 57–68.

Meier, O., Greber, U.F., 2003. Adenovirus endocytosis. J. Gene Med. 5 (6), 451–462.

Rink, J., Ghigo, E., Kalaidzidis, Y., Zerial, M., 2005. Rab conversion as a mechanism of progression from early to late endosomes. Cell 122 (5), 735–749.

Ponti, A., Vallotton, P., Salmon, W.C., Waterman-Storer, C.M., Danuser, G., 2003. Computational analysis of F-actin turnover in cortical actin meshworks using fluorescent speckle microscopy. Biophys. J. 84 (5), 3336–3352.

Sbalzarini, I.F., Koumoutsakos, P., 2005. Feature point tracking and trajectory analysis for video imaging in cell biology. J. Struct. Biol. 151, 182–195.

Qian, H., Sheetz, M.P., Elson, E.L., 1991. Single-particle tracking. Analysis of diffusion and flow in two-dimensional systems. Biophys. J. 60 (4), 910–921.

Ferrari, R., Manfroi, A.J., Young, W.R., 2001. Strongly and weakly self-similar diffusion. Physica D 154, 111–137.

Ewers, H., Smith, A.E., Sbalzarini, I.F., Lilie, H., Koumoutsakos, P., Helenius, A., 2005. Single-particle tracking of murine polyoma virus-like particles on live cells and artificial membranes. Proc. Natl. Acad. Sci. USA 102 (42), 15110–15115.

Kinder, M., Brauer, W., 1993. Classification of trajectories—extracting invariants with a neural network. Neural Network 6 (7), 1011–1017.

Owens, J., Hunter, A. Application of the self-organizing map to trajectory classification. In: Third IEEE International Workshop on Visual Surveillance (VS'2000), pp. 77–83.

Fraile, R. Maybank, S.J. Vehicle trajectory approximation and classification, Report, University of Reading, Dept. of Computer Science.

Chen, Y., Yang, B., Jacobson, K., 2004. Transient confinement zones—a type of lipid raft?. Lipids 39 (11) 1115–1119.

Kusumi, A., Nakada, C., Ritchie, K., Murase, K., Suzuki, K., Murakoshi, H., Kasai, R.S., Kondo, J., Fujiwara, T., 2005. Paradigm shift of the plasma-membrane concept from the two-dimensional continuum fluid to the partitioned fluid: high-speed single-molecule tracking of membrane molecules. Annu. Rev. Bioph. Biom. Struct. 34, 351–378.

Huet, S., Karatekin, E., Tran, V.S., Fanget, I., Cribier, S., Henry, J.-P., 2006. Analysis of transient behavior in complex trajectories: application to secretory vesicle dynamics. Biophys. J. 91, 3542–3559.

Horwitz, M.S., 2001. Fields Virology. Raven Press, Philadelphia, Pa (Ch. Adenoviruses, pp. 2301–2326).

Medeiros, N.A., Burnette, D.T., Forscher, P., 2006. Myosin II functions in actin-bundle turnover in neuronal growth cones. Nat. Cell Biol. 8 (3), 215–226.

Lehmann, M.J., Sherer, N.M., Marks, C.B., Pypaert, M., Mothes, W., 2005. Actin- and myosin-driven movement of viruses along filopodia precedes their entry into cells. J. Cell Biol. 170 (2), 317–325.

Vale, R.D., 2003. The molecular motor toolbox for intracellular transport. Cell 112 (4), 467–480.

Suomalainen, M., Nakano, M.Y., Keller, S., Boucke, K., Stidwill, R.P., Greber, U.F., 1999. Microtubule-dependent plus- and minus end-directed motilities are competing processes for nuclear targeting of adenovirus. J. Cell Biol. 144 (4), 657–672.

Schölkopf, B., Smola, A.J., 2002. Learning with Kernels. Support vector machines, regularization, optimization and beyond. MIT Press, Cambridge, Massachusetts.

Einstein, A., 1905. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Annalen der Physik 17, 549–560.

Saxton, M., 2001. Anomalous sub-diffusion in fluorescence photobleaching recovery: a monte carlo study. Biophys. J. 81, 2226–2240.

Weiss, M., Hashimoto, H., Nilsson, T., 2003. Anomalous protein diffusion in living cells as seen by fluorescence correlation spectroscopy. Biophys. J. 84, 4043–4052.

Berry, M., Lewis, Z., 1980. On the Weierstrass-Mandelbrot fractal function. P. Roy. Soc. Lond. A Mat. 370 (1743), 459–484.

Desai, A., Mitchison, T.J., 1997. Microtubule polymerization dynamics. Annu. Rev. Cell Dev. Biol. 13, 83–117.

Dogterom, M., Kerssemakers, J., Romet-Lemonne, G., Janson, M., 2005. Force generation by dynamic microtubules. Curr. Opin. Cell Biol. 17, 67–74.

Hastie, T., Tibshirani, R., Friedman, J., 2003. The Elements of Statistical Learning, 1st ed. Springer.

Luby-Phelps, K., 2000. Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area. Int. Rev. Cytol. 192, 189–221.

Sodeik, B., 2000. Mechanisms of viral transport in the cytoplasm. Trends Microbiol. 8 (10), 465–472.

ORourke, N.A., Dailey, M.E., Smith, S.J., McConnell, S.K., 1992. Diverse migratory pathways in the developing cerebral–cortex. Science 258 (5080), 299–302.

Soo, F.S., Theriot, J.A., 2005. Large-scale quantitative analysis of sources of variation in the actin polymerization-based movement of Listeria monocytogenes. Biophys. J. 89, 703–723.

Wolf, H., Wehner, R., 2000. Pinpointing food sources: Olfactory and anemotactic orientation in desert ants, Cataglyphis fortis. J. Exp. Biol. 203 (5), 857–868.

Croxall, J., Silk, J., 2005. Global circumnavigations: tracking year-round ranges of non-breeding albatrosses. Science 307, 249–250.