Identifying Virus-Cell Fusion in Two-Channel Fluorescence Microscopy Image Sequences Based on a Layered Probabilistic Approach

W. J. Godinez*, Student Member, IEEE, M. Lampe, P. Koch, R. Eils, B. Müller, and K. Rohr, Member, IEEE

Abstract—The entry process of virus particles into cells is decisive for infection. In this work, we investigate fusion of virus particles with the cell membrane via time-lapse fluorescence microscopy. To automatically identify fusion for single particles based on their intensity over time, we have developed a layered probabilistic approach. The approach decomposes the action of a single particle into three abstractions: the intensity over time, the underlying temporal intensity model, as well as a high level behavior. Each abstraction corresponds to a layer and these layers are represented via stochastic hybrid systems and Hidden Markov Models. We use a maxbelief strategy to efficiently combine both representations. To compute estimates for the abstractions we use a hybrid particle filter and the Viterbi algorithm. Based on synthetic image sequences, we characterize the performance of the approach as a function of the image noise. We also characterize the performance as a function of the tracking error. We have also successfully applied the approach to real image sequences displaying pseudotyped HIV-1 particles in contact with host cells and compared the experimental results with ground truth obtained by manual analysis.

Index Terms—Biomedical imaging, microscopy images, tracking, virus particles, behavior identification.

I. INTRODUCTION

INFECtIONS caused by viruses represent a serious threat to human life. The prejudicial potential of viral infections is highlighted by pandemic outbreaks of viruses such as the H1N1 influenza A virus (e.g., [1]), the human-immunodeficiency virus type 1 (HIV-1) (e.g., [2]), as well as the hepatitis C virus (e.g., [3]). Entry of viruses into cells is crucial for infection. In general, an enveloped virus particle, such as HIV-1, can enter a cell via two alternative mechanisms: endocytosis or fusion at the plasma membrane ([4], [5]). In the former case, the cell internalizes the virus particle by engulfing it into an intracellular vesicle. In the latter case, the virus particle directly fuses with the plasma membrane. However, the dynamics of these entry processes are not well understood. Detailed insight into the fusion process results in a better understanding of the infectivity of the viruses, which in turn can lead to the development of more effective antiviral drugs. In this work, we investigate the fusion mechanisms of pseudotyped human immunodeficiency virus type 1 (HIV-1) particles with the cell membrane using live-cell microscopy as well as double fluorescence labeling strategies ([6]–[8]). Concretely, each virus particle is tagged with two different fluorescent labels: one label is attached to the outer shell of the particle while the other is attached to its inner part. This labeling technique results in two-channel fluorescence microscopy image sequences. Upon fusion with the cell membrane, the particle’s outer shell is disrupted and the inner part is released into the cell. Fusion should lead to the disappearance of the fluorescent label attached to the outer shell because this label is diluted into the cell membrane, and the label attached to the inner part should remain visible. Hence, fusion can be defined as a transient behavior that is implicitly described by changes in the fluorescence intensities of a single particle. Such a behavior can be observed by tracking single fluorescent particles (e.g., [9]–[17]), which yields the individual temporal statistics (e.g., the intensity over time) required for determining the current behavior of each particle. To draw statistically sound conclusions about the fusion behavior of viruses, a large number of individual particles must be analyzed. Thus, approaches that automatically identify fusion based on the appearance of individual virus particles are required.

However, identifying behaviors based on the appearance of virus particles is challenging. Because of the small size of the virus particles (ca. 140 nm in the case of HIV-1), conventional fluorescence microscopy cannot spatially resolve these objects. Instead, virus particles are displayed as small ‘spots’ with hardly any salient visual features (e.g., texture, shape) that could aid the identification of virus behavior. Additionally, autofluorescent proteins with relatively weak signal intensity instead of synthetic fluorophores have to be used, and comparatively low laser power has to be used for long-term observation of live cells. These factors lead to a relatively low fluorescence intensity. The intensity of the particles is further subjected to distortions caused by other autofluorescent structures, such as cells. Photobleaching is likewise problematic since it reduces the fluorescence intensity of the particles over time. Also, shot noise introduces a random fluctuation on the particle’s intensity. In addition, movement of the particles into
out-of-focus planes also alter their appearance (intensity as well as size). Therefore, automatic approaches must cope with data that is both limited and distorted.

A. Previous Work

Behaviors of fluorescent particles, such as fusion, are often reflected as fluctuations of the temporal intensity statistics of individual particles. Changes in the intensity over time may be detected at the image level by computing the difference image between two consecutive images using single pixels (e.g., [18]–[20]) or image regions (e.g., [21]). Since changes in the intensity may also arise from other phenomena (e.g., motion of the particles, image noise), approaches based on image differencing may not discriminate accurately between changes in the intensity arising from behaviors of interest (e.g., fusion) and changes originating from other phenomena. Also, these approaches are typically not applicable for detecting behaviors that entail changes in temporal statistics other than the intensity (e.g., size of a particle). By tracking the fluorescent particles and thereby obtaining the temporal statistics of each object, the changes can be detected at the object level. For example, within the context of neurobiology, the fusion of a fluorescent neurotransmitter vesicle with the cell membrane is detected by fitting a linear model to the curve of estimated radii over time of the fluorescent particle ([22]). A statistical test on the fitted slope parameter, which reflects the rate of change in the particle’s radii, determines whether fusion occurred. Such a global approach, while taking advantage of all temporal data, assumes a constant behavior and is not able to identify potentially disparate behaviors over time. In comparison, local approaches identify behaviors of interest at each time step by taking into account the statistics of few consecutive time steps.

Local approaches that detect fusion based on the intensity over time of an individual particle typically exploit the fact that this behavior entails a rapid change in the intensity. For example, [23] used a derivative-based approach, where fusion of an influenza virus particle with the cell membrane is represented by large values of the first derivative of the particle’s intensity with respect to time. To numerically estimate the derivatives typically small temporal neighborhoods are incorporated and therefore these approaches are susceptible to noise. In [24], [25] the rate of change in intensity has been used for detecting fusion of glucose transporter 4 (GLUT4) vesicles with the cell membrane. Local approaches do not take into account results from previous time steps for analyzing subsequent time steps, i.e., the temporal coherence of the particle’s behavior is not exploited. Approaches based on Hidden Markov Models (HMMs) address this shortcoming by specifying local transition probabilities between different behaviors, thereby encouraging plausible behavior sequences. For instance, [26], [27] as well as [28] employ HMMs for analyzing the fluorescence intensity of single molecules. However, HMMs estimate discrete variables (states) only and typically represent the temporal statistics using piecewise constant functions. Approaches based on stochastic hybrid systems overcome these limitations by estimating both the discrete behavior as well as the continuous temporal statistics.

Stochastic hybrid systems have first been introduced in automatic control applications (e.g., [29]), and have been recently revisited for behavior identification in the field of computer vision. Applications include identification of behaviors based on face expressions (e.g., [30], [31]), as well as on human gaits and pose (e.g., [32], [33]). Other applications include the identification of the behavior of bees (e.g., [34], [35]). In biological imaging, hybrid stochastic systems have been used for estimating the position of fluorescent particles (e.g., [10], [12]) or cells (e.g., [36]), however, in these approaches the hybrid stochastic system does not exploit the intensity information of the biological objects and the estimated discrete variables are not used for behavior identification. In general, estimation of the hybrid states is often carried out within a Bayesian framework where the aim is to compute a posterior distribution on the hybrid states given the observed temporal statistics of the object under consideration. This task is challenging since the optimal solution entails an exponential computational effort with respect to time. Therefore, approximate algorithms have been proposed. For example, [35] suggested a greedy approach that involves first approximating the posterior distribution of the continuous variables followed by calculating maximum likelihood (ML) estimates for the discrete behavior variables. While this approach significantly reduces the complexity of the estimation task, the greedy ML estimates do not consider the validity of transitions between behaviors. Another approach uses an approximation of the posterior via a mixture of Gaussianians (e.g., [32]). However, this approach is only applicable to linear Gaussian models. In [34], a data-driven Markov Chain Monte Carlo (MCMC) method is used to approximate the posterior. There, the MCMC proposal distribution is derived from the likelihood of the observed temporal statistics as well as from the transition probabilities between different behaviors. While the solution space is efficiently explored, the approach requires a relatively large amount of training data that should include a label for each behavior at each time step. Additionally, MCMC methods do not typically support sequential analysis of the observed temporal statistics, which limits their applicability to off-line inference tasks.

Particle filters (e.g., [37]) provide a sequential alternative to MCMC methods. Within computer vision, the application of particle filters for carrying out inference in stochastic hybrid systems is first described in [38]. Because of its applicability to non-linear and non-Gaussian models as well as its recognition-by-synthesis approach, such a hybrid particle filter has seen increasing interest for identifying behaviors. For example, [39] represent behaviors (e.g., facial expressions) as temporal trajectories within a certain space (e.g., a space defined by the basis of optical flow fields). The temporal trajectories are sequentially matched to a temporal neighborhood of the observed temporal statistics via a hybrid particle filter, where the temporal neighborhood is defined by a temporal window. The length of this window influences the results and determining the optimal length is not trivial. In certain applications, the transition probabilities for certain behaviors may be low, which entails that unlikely behaviors may be supported by very few hybrid samples. The lack of support in certain regions of the discrete space reduces the accuracy
of the approximated hybrid posterior. One strategy to cope with this issue involves carrying out importance sampling on the discrete variable only, whereby the discrete space is explored more effectively by using an importance transition matrix \((40)\). However, performing importance sampling on the discrete variable only may generally decrease the accuracy of the estimates for the continuous variables. Another strategy to cope with the lack of support for unlikely behaviors is to increase the number of samples. This straightforward strategy works well at the expense of an increase in the computational cost, which scales linearly with respect to the number of samples. To compensate for the high computational cost, some approaches (e.g., \([41]\)) assume that the behavior remains fixed over time. Such an approach reduces significantly the extent of the solution space but it is not applicable to the task of identifying disparate behaviors over time. Other schemes (e.g., \([31]\)) assume no direct dependence between the discrete variable and the continuous variables, but such an assumption may not hold in several applications.

An additional issue with the hybrid particle filter is that the estimates for the variables are computed sequentially using only the temporal statistics found up to the current time step. Using all available temporal statistics may increase the accuracy of the computed estimates. A smoothing algorithm based on particle filters for computing discrete and continuous variables based on all available temporal statistics has been presented in \([42]\). However, the computational cost of such a smoothing scheme is quadratic with respect to the number of samples. Most schemes based on the hybrid particle filter deal with one object only, so smoothing may be applicable in those cases. However, in cases where multiple objects are present over several time steps, such a smoothing scheme may be impractical.

**B. Our Approach**

In this paper, based on our previous work in \([43]\), we introduce an automatic approach grounded within the theory of Bayesian estimation for identifying fusion of double labeled enveloped virus particles with the cell membrane in twocolor fluorescence microscopy images. Our approach adopts a layered architecture that decomposes the actions of a single virus particle into three abstractions (viz., the intensity, the underlying temporal intensity model, and the behavior). The three abstractions are represented by the different layers, which are in turn described using a stochastic hybrid system and an HMM. The two representations are combined via a maxbelief strategy (e.g., \([44]\)). To estimate the variables of the layers we use a hybrid particle filter (e.g., \([45]\)) as well as the Viterbi algorithm. Our layered approach entails several advantages: first, the combination of stochastic hybrid systems and HMMs offer an improved modeling capability. Second, the maxbelief strategy is straightforward and introduces no additional computational overhead. Third, the modularity of such a layered approach endows our scheme with efficiency, since the dimensionality of the state space of the different layers is low. Because of the low dimensionality, a relatively low number of samples is sufficient to ensure a good representation of the hybrid posterior. Additionally, the layers can also be adjusted; thus the approach is flexible and can be straightforwardly adapted for identifying other behaviors. We also introduce temporal intensity models to describe the local fluctuations of the intensity of single virus particles. The models are defined as auto-regressive (AR) processes and thus allow an intuitive interpretation of the corresponding process parameters. In contrast to global approaches (e.g., \([22]\)), our approach accounts for changes in the behavior over time. In contrast to local approaches (e.g., \([23]\)), our approach defines local transition probabilities to enforce coherent sequences of temporal intensity models as well as of behaviors. In contrast to single-layer HMMs, the intensity statistics are more accurately modeled using temporal intensity models. Compared to our previous work in \([43]\), where only the intensity statistics from a single channel were used, here we introduce a hierarchical approach that exploits efficiently the intensity statistics of both channels. Note that our approach for behavior identification builds upon a virus tracking approach (e.g., \([16]\)) which is performed prior to applying our approach. Based on synthetic images, we describe the performance of the layered approach as a function of the image noise. We also characterize the performance as a function of the tracking error. In addition, we have successfully applied the approach to real images displaying HIV-1 particles. To the best of our knowledge, this is the first time that a layered probabilistic approach has been used for identifying behaviors of fluorescent particles in multi-channel microscopy image sequences.

This paper is organized as follows. In Section II, we describe the layered probabilistic approach for identifying fusion. Section III presents the models describing the behavior of virus particles. The experimental results are described in Section IV and then conclusions are presented in Section V.

**II. THREE-LAYER PROBABILISTIC APPROACH FOR FUSION IDENTIFICATION**

**A. Overview of the Approach**

Our layered approach consists of three layers: The first layer corresponds to the intensity of a particle, and the second layer represents the temporal intensity model of the particle. The third and topmost layer models the behavior (including fusion) of the particle. The intensity is described via autoregressive (AR) processes while the temporal intensity models follow a first-order Markov chain. The two layers are jointly modeled using a stochastic hybrid system. The hybrid system describes the intensity and temporal intensity models of one channel and is thus used separately on the temporal intensity statistics of each channel of the two-channel image data. To take into account the information from both channels, the identified temporal intensity models in the individual channels are fed to the layer modeling the behavior. The behavior is described via an HMM. By using the Viterbi algorithm we obtain the most probable sequence of behaviors for a particle.

**B. Stochastic Hybrid Systems of Virus Intensity: Bayesian Framework**

In our approach for identifying fusion of virus particles with the cell membrane we represent the intensity of a virus particle
via a set of temporal intensity models that take the form of non-linear autoregressive processes. At a given time step it is assumed that only one temporal intensity model specifies the observed intensity of the particle. The aim is to estimate the intensity as well as the current temporal intensity model from a given sequence of observed intensities. We formulate this task as a sequential hybrid estimation problem. Within the theory of stochastic hybrid systems (e.g., [46], [47]), it is assumed that a variable of interest of a system can be characterized via a set $A$ of $N_A$ predefined system models. At time step $t$, the variable is both governed by a certain model $\alpha_t \in A$, and manifested via a state vector $x_t$ (in our case, the intensity of a particle), which in turn is reflected by a noisy measurement $y_t$ (in our case, the intensity of a particle obtained from the image data). The goal is to estimate $(x_t, \alpha_t)$ given a sequence of measurements $y_{1:t}$. A Bayesian approach involves computing the posterior densities $p(x_t, \alpha_t | y_{1:t})$, which can be factorized as:

$$p(x_t, \alpha_t | y_{1:t}) = p(x_t | \alpha_t, y_{1:t}) p(\alpha_t | y_{1:t}).$$

(1)

This implicates computing the model-conditioned posterior densities $p(x_t | \alpha_t, y_{1:t})$ as well as the posterior model probabilities $p(\alpha_t | y_{1:t})$. At time step $t - 1$, all $N_A$ model-conditioned posteriors are maintained. Since the evolution of the model $\alpha_{t-1}$ is represented using a Markov chain associated with a transition matrix $\Pi$, each model $\alpha_{t-1} = j$ may ‘branch’ into any other model $\alpha_t = i$ at the next time step $t$ with probability $\pi_{ij}$. The optimal solution involves computing each possible model history ([29], [46]), i.e., computing the full model history tree, which leads to an exponential increase in computational effort w.r.t. time. Suboptimal strategies aim at maintaining a constant computational load over time. One such strategy (e.g., [45], [47], [48]) with application to biological particles (e.g., [10], [12]) involves first computing the prior model probabilities $P(\alpha_t | y_{1:t-1})$ via the Chapman-Kolmogorov equation for the Markov chain underlying the model evolution:

$$P(\alpha_t | y_{1:t-1}) = \sum_{\alpha_{t-1}=1}^{N_A} P(\alpha_{t-1} | y_{1:t-1}) P(\alpha_t | \alpha_{t-1}),$$

(2)

The branches entailed by the evolution of the Markov chain are ‘mixed’ into $N_A$ model-conditioned prior densities $p(x_{t-1} | \alpha_t, y_{1:t-1})$. Assuming that $x_{t-1}$ is independent of $\alpha_t$ given $\alpha_{t-1}$, the mixing of the densities amounts to a weighted sum of the model-conditioned posteriors:

$$p(x_{t-1} | \alpha_t, y_{1:t-1}) = \sum_{\alpha_{t-1}=1}^{N_A} p(x_{t-1} | \alpha_{t-1}, y_{1:t-1}) P(\alpha_{t-1} | \alpha_t, y_{1:t-1}).$$

(3)

Here the weights are obtained by applying Bayes’ theorem:

$$P(\alpha_{t-1} | \alpha_t, y_{1:t-1}) = \frac{P(\alpha_{t-1} | y_{1:t-1}) P(\alpha_t | \alpha_{t-1})}{P(\alpha_t | y_{1:t-1})}.$$  

(4)

Each of these $N_A$ mixture prior densities can now be employed within a model-matched filtering framework. Concretely, the model-conditioned mixture priors in (3) are propagated over time by applying the corresponding model-conditioned dynamical model $p(x_t | x_{t-1}, \alpha_t)$:

$$p(x_t | \alpha_t, y_{1:t-1}) = \int p(x_t | x_{t-1}, \alpha_t) p(x_{t-1} | \alpha_t, y_{1:t-1}) dx_{t-1}. \quad (5)$$

Using a model-conditioned measurement model $p(y_t | x_t, \alpha_t)$, an update of the predicted densities $p(x_t | \alpha_t, y_{1:t-1})$ is obtained via Bayes’ theorem:

$$p(x_t | \alpha_t, y_{1:t}) \propto p(y_t | x_t, \alpha_t) p(x_t | \alpha_t, y_{1:t-1}). \quad (6)$$

Likewise the predicted probabilities $P(\alpha_t | y_{1:t-1})$ are updated using Bayes’ theorem:

$$P(\alpha_t | y_{1:t}) = \frac{p(y_t | \alpha_t, y_{1:t-1}) P(\alpha_t | y_{1:t-1})}{\sum_{\alpha_t} p(y_t | \alpha_t, y_{1:t-1}) P(\alpha_t | y_{1:t-1})}. \quad (7)$$

where the denominator acts as a normalization factor, and where, by the law of total probability and assuming that the measurements are independent of each other, the likelihood $p(y_t | \alpha_t, y_{1:t-1})$ is defined as:

$$p(y_t | \alpha_t, y_{1:t-1}) = \int p(y_t | x_t, \alpha_t) p(x_t | \alpha_t, y_{1:t-1}) dx_t. \quad (8)$$

This strategy consists of $N_A + 1$ estimation processes: $N_A$ model-matched filters (see (5) and (6)) as well as an estimator for the model (see (2) and (7)). Note that the $N_A$ model-matched filters interact with the model estimator via (3) as well as via (7). However, handling the $N_A$ mixture priors in (3) that are used as input for the model-matched filters could be cumbersome, since these priors can have a complex form. Algorithms adopting this strategy therefore reduce the complexity of the priors by pruning or merging the components of the involved mixture densities. For example, the Interacting Multiple Model (IMM) algorithm ([48], [49]) assumes that the models are linear and Gaussian and thus each prior amounts to a Gaussian mixture. The IMM converts this Gaussian mixture to a single Gaussian density of matched first and second moments that is amenable to analytical techniques based on the Kalman filter. In cases where the models are non-linear and/or non-Gaussian, exact calculations of the above (multivariate) relations are in general not possible. However, one may obtain a sound numerical approximation of these distributions via particle filters. An approach combining the IMM algorithm with particle filters has been introduced in [47] (see also [12]). Whereas in the standard IMM each prior (3) is approximated by a single Gaussian density, in [47] each prior is more accurately represented by a non-parametric mixture density with $N_A$ components, where each component is approximated by a particle filter with a fixed number of samples. In comparison to that, in our work we resolve the recursions using a hybrid particle filter that approximates the priors by directly simulating the dynamical models governing the evolution of the hybrid system over time (i.e., each sample represents an instance of the hybrid system at a given time step, cf. [45]). We use the hybrid particle filter to estimate both the discrete and continuous variables, and we do not assume a Gaussian density for the likelihood of each model. We use
samples for all $N_A$ models, and the number of samples for each model varies over time. Using a large number of samples $N_A$ relative to the dimension of the state space allows our approach to cope well with relatively small values for the transition probabilities, which represent rare events.

C. Stochastic Hybrid Systems of Virus Intensity: Implementation via Hybrid Particle Filter

Based on the principle of importance sampling, a set $\{x_k^i; w_k^i\}_{k=1}^{N_s}$ of $N_s$ random samples $x_k^i$ (the ‘particles’) associated with importance weights $w_k^i$ is assumed to approximately represent an arbitrary distribution $p(x)$ in the sense that selecting a certain $x_k^i$ with probability proportional to $w_k^i$ amounts to drawing a random sample from $p(x)$. The particle filter (e.g., [37, 50]) is a numerical scheme for recursively constructing such a set $\{x_k^i; w_k^i\}_{k=1}^{N_s}$ for approximating the posterior distribution $p(x_t | y_{1:t})$, which is defined by stochastic propagation and Bayes’ theorem as:

$$p(x_t | y_{1:t}) \propto p(y_t | x_t) \int p(x_t | x_{t-1}) p(x_{t-1} | y_{1:t-1}) dx_{t-1}.$$  
(9)

The weighted approximation of the posterior obtained by the particle filter is thus given as:

$$p(x_t | y_{1:t}) \approx \sum_{k=1}^{N_s} w_k^i \delta(x_t - x_k^i),$$  
(10)

where $\delta(\cdot)$ denotes the Dirac delta measure and $\sum_{k=1}^{N_s} w_k^i = 1$. Briefly, the particle filter resolves the (single model) Bayesian recursion defined in (9) in three steps. Starting with the set $\{x_k^i; w_k^i\}_{k=1}^{N_s}$, the filter first generates each predicted sample $x_k^i$ with a random draw from the dynamical model $p(x_t | x_{t-1})$. Second, these predicted samples are re-weighted according to $w_k^i \propto p(y_t | x_k^i)$. The weights are normalized so that they sum to unity. Finally, to select the most representative samples, the filter resamples $N_s$ times from the set of weighted samples $\{x_k^i; w_k^i\}_{k=1}^{N_s}$. Note that within the context of particle filters, the term ‘particle’ refers to a sample drawn from a certain distribution. This term has a different meaning than the term ‘virus particle’ (referring to an observed virus particle in an image), for which also the abbreviated term ‘particle’ is used. To avoid confusion, in the following, we use the term ‘sample’ to refer to a ‘particle’ of a particle filter. We use the word ‘particle’ to refer to a virus particle.

One approach for extending the particle filter to stochastic hybrid systems consists of augmenting each $k$-th sample with a model index $\alpha_k^i$ thereby obtaining a set of hybrid samples $\{x_k^i, \alpha_k^i; w_k^i\}_{k=1}^{N_s}$ (e.g., [45]). Based on this set of hybrid samples, the posterior in (1) can be represented by the following weighted approximation (151):

$$p(x_t, \alpha_t | y_{1:t}) \approx \sum_{k=1}^{N_s} w_k^i \delta(x_t - x_k^i) \Pi_{\alpha_k^i}(\alpha_t).$$  
(11)

Here $\Pi(\cdot)$ is the indicator function, where $\Pi_{\alpha}(\cdot) = 1$ if $\alpha = \alpha$, else $\Pi_{\alpha}(\cdot) = 0$. At time step $t$ the hybrid particle filter maintains $N_A$ model-conditioned posterior densities $p(x_t | \alpha_t = i, y_{1:t})$, where $i \in A$. Each $i$-th model-conditioned posterior is thus approximated via the set of samples $\{x_t^i, \alpha_t^i = i\}$:

$$p(x_t | \alpha_t = i, y_{1:t}) \approx \sum_{k \in \Upsilon_i} \tilde{w}_k^i \delta(x_t - x_k^i),$$  
(12)

where $\Upsilon_i = \{k | x_t^i, \alpha_t^i = i\}$ and $\tilde{w}_k^i = \frac{w_k^i}{\sum_{k \in \Upsilon_i} w_k^i}$. Similarly, the posterior model probabilities $P(\alpha_t = i | y_{1:t})$ are approximated by the proportion of samples supporting model $i$:

$$P(\alpha_t = i | y_{1:t}) \approx \frac{| \{k | x_t^i, \alpha_t^i = i\} |}{N_s},$$  
(13)

where $\cdot$ denotes the set size operator. The hybrid particle filter recursively constructs a set of hybrid samples as follows. At time step $t - 1$, the predicted model index $\alpha_{t-1}^i$ is generated by drawing a random sample from $P(\alpha_{t-1} | y_{1:t-1})$, which is defined by the $j$-th column of the transition matrix $\Pi$. Samples that branch from model $\alpha_{t-1}^i = j$ into model $\alpha_t^i = i$ represent approximately the branched priors $p(x_{t-1} | \alpha_{t-1}^i = j, \alpha_t^i = i, y_{1:t-1})$. Following the mixing strategy as expressed by (3), the model-conditioned prior densities $P(x_t | \alpha_t = i, y_{1:t-1})$ are supported by the set of samples with predicted model index $\alpha_t^i = i$. The particle-based representation of each $i$-th model-conditioned prior is suitable for the sequential importance sampling-resampling scheme of the standard particle filter. Specifically, a prediction for $x_t^i$ is obtained by drawing a random sample from the dynamical model $p(x_t | \alpha_{t-1}^i = i, y_{1:t-1})$. The resampling step is finally applied to the entire set of samples to obtain a set $\{x_t^i, \alpha_t^i; w_t^i\}_{k=1}^{N_s}$ of $N_s$ hybrid samples. At the end of each iteration, a maximum a posteriori (MAP) estimate for the model $\alpha_t$ may be obtained from the model posterior probabilities (e.g., [38]):

$$\hat{\alpha}_t = \arg \max_i P(\alpha_t = i | y_{1:t}).$$  
(14)

An estimate for $x_t$ may be computed as the weighted mean over all $N_s$ samples:

$$x_t = \sum_{k=1}^{N_s} w_t^i x_t^i.$$  
(15)

Alternatively, one may obtain the mean estimate of the model-conditioned posterior corresponding to the MAP estimate $\hat{\alpha}_t$:

$$x_t = \sum_{k \in \hat{T}} \tilde{w}_k^i x_t^i,$$  
(16)

where $\hat{T} = \{k | x_t^i, \alpha_t^i = \alpha_t\}$ and $\tilde{w}_k^i = \frac{w_k^i}{\sum_{l \in \hat{T}} w_l^i}$.

D. Hidden Markov Model of Virus Behavior

In our experimental setup, fusion of a pseudotyped HIV-1 particle with the cell membrane in two-channel microscopy image sequences is characterized by both the loss of the
fluorescent label attached to the outer shell in one channel as well as the preservation of the label attached to the particle’s inner part in the other channel. Thus, the intensity as well as the underlying temporal intensity models of the two channels must be considered when automatically identifying individual fusion events. Using stochastic hybrid systems for identifying the temporal intensity models of both channels entails increasing the dimensionality of the state vector $x_t$ as well as increasing the number of models $N_A$. In the worst case, one would need to define $N_A^2$ models accounting for the joint dynamics of the intensities of both channels. This would require building a transition matrix $\Pi$ of size $N_A^2 \times N_A^2$, which could be impractical to define. Because of the scarce amount of available training data, learning the transition matrix may lead to inaccurate estimates of the transition probabilities. Additionally, increasing the number of models $N_A$ requires a significant increase in the number of samples for the hybrid particle filter to obtain a comparable performance. Instead of using this monolithic strategy, we adopt the following hierarchical solution. First, we use hybrid particle filters for the intensities measured on each channel separately. This yields a sequence $\alpha_{1:T}$ of $T$ temporal intensity models per channel $c$ that correspond to the MAP model estimates, i.e., the models with the highest belief obtained at each time step $t$. The sequence of vectors $\alpha_{1:T}$, which comprises the inferred models of both channels, is fed to a Hidden Markov Model (HMM) (e.g., [52]) that maps these sequences to a sequence of behaviors $\beta_{1:T}$, where $\beta$ represents a certain behavior from a set $B$ of $N_B$ predefined behaviors. Concretely, a Markov chain with transition matrix $\Phi$ is assumed to underlie the sequence of behaviors $\beta_{1:T}$. Using the HMM formalism, one can define a posterior distribution over the sequence of behaviors $\beta_{1:T}$ given a sequence of observed intensity models $\alpha_{1:T}$ as follows:

$$P(\beta_{1:T} | \alpha_{1:T}) \propto \prod_{t=1}^{T} P(\beta_t | \beta_{t-1}) P(\alpha_t | \beta_t),$$

(17)

where $P(\beta_t = i | \beta_{t-1} = j)$ is given by the corresponding element $\phi_{ij}$ of the transition matrix $\Phi$. $P(\alpha_t | \beta_t)$ denotes the measurement model defined by the HMM. Our aim is to determine the MAP sequence of behaviors:

$$\hat{\beta}_{1:T} = \arg \max_{\beta_{1:T}} P(\beta_{1:T} | \alpha_{1:T}).$$

(18)

Note that (18) can be efficiently solved using the Viterbi algorithm (e.g., [53]). Thus, our approach represents the behavior $\beta$, the temporal intensity model $\alpha$, as well as the intensity statistics $x$ via three stacked Markov processes that amount to a layered probabilistic approach where a maxbelief strategy ([44]) is used for linking the two top layers (namely behaviors and intensity models) in a bottom-up fashion.

III. MODEL DEFINITIONS

A. Temporal Intensity Models

In this section, we define explicit representations of the temporal intensity models underlying the observed intensities. The models must account for the various phenomena exhibited by the observed intensity values. For example, because of photobleaching, the intensity exhibits a slow downward trend. In comparison, fusion with the cell membrane is characterized by a steep decrease in the intensity of one channel. The models should account for the different magnitudes of the change in intensity. In our case, the state vector $x$ is given by the scalar value corresponding to the intensity $I$ (which refers to the object intensity plus the background intensity). We define $N_A = 3$ intensity models: constant intensity ($CI, \alpha = 0$), positive intensity change ($PIC, \alpha = 1$), and negative intensity change ($NIC, \alpha = 2$). The corresponding dynamical models $p(x_t | x_{t-1}, \alpha_t)$ are described via first-order auto-regressive processes on the state vector $x_t$, which take the following form:

$$x_t = f_{\alpha_t}(x_{t-1}, v_{\alpha_t}),$$

(19)

where $v_{\alpha_t}$ is a noise variable sampled from a zero-mean Gaussian distribution with variance $Q_{\alpha_t}$. To define the function $f(.)$ in (19), the CI model ($\alpha = 0$) adopts the following relation:

$$f_{\alpha_t}(x_{t-1}, v_{\alpha_t}) = I_{t-1} + v_{\alpha_t}$$

(20)

For this model, we define $Q_{\alpha_t} = q_{ci}$, where $q_{ci}$ is the variance of the deviation in the intensity over a time interval. For the PIC model ($\alpha = 1$), the function $f(.)$ in (19) is given as:

$$f_{\alpha_t}(x_{t-1}, v_{\alpha_t}) = I_{t-1} + |v_{\alpha_t}|$$

(21)

and $Q_{\alpha_t} = q_{pic}$. Likewise, the NIC model ($\alpha = 2$) defines $f(.)$ in (19) as:

$$f_{\alpha_t}(x_{t-1}, v_{\alpha_t}) = I_{t-1} - |v_{\alpha_t}|$$

(22)

and $Q_{\alpha_t} = q_{nic}$. Analogous to the CI model, $q_{pic}$ and $q_{nic}$ regulate the deviation for the intensity variable over time. With a relatively small value for $q_{ci}$, the CI model captures smooth changes in the intensity introduced by photobleaching as well as other auto-fluorescent structures. By using a relatively large value for $q_{nic}$, the NIC model represents sharp decreases in the intensity, which in our application correspond to the specific behavior of fusion. The PIC model is added for completeness. In addition to the model-conditioned dynamical model, each temporal intensity model $\alpha$ could in principle define a model-conditioned measurement model $p(y_t | x_t, \alpha_t)$. In our case all models share a common measurement model, i.e., $p(y_t | x_t, \alpha_t) = p(y_t | x_t)$. The proposed measurement model quantifies the probability that the predicted intensity $I$ matches the particle’s mean intensity $y$ measured from the image:

$$p(y | x) \propto \exp \left(- \frac{D(y, I)^2}{2\sigma_n^2} \right),$$

(23)

where $D(.)$ is the Euclidean distance and $\sigma_n$ describes the expected level of noise.

In our application, virus particles typically exhibit a constant behavior and fusion is rare. In other words, we assume that once the intensity follows a certain temporal model, the intensity will not jump between models but instead will adhere to a model for a certain time period. The entries $\pi_{ij}$ of the transition matrix $\Pi$ reflect these considerations. Because of this model steadiness, the matrix $\Pi$ is strongly diagonal, i.e.,
transition probabilities were kept fixed for all our experiments. For stability reasons, we also assume that the intensity cannot switch directly from a PIC model to a NIC model, or vice-versa. While the transition matrix could be learned from training data, we manually set the entries for the following reasons. First, large amounts of training data would be required to obtain sound estimates for the transition probabilities, especially since fusion of HIV-1 particles with the cell membrane is relatively rare ([54]), and thus the number of exemplary periods of intensity changes (e.g., decrease of intensity) is inherently low. In addition, this training data would need to include one model label for each time step. Such a labeling would be quite tedious and time-consuming to obtain. Second, manually setting the values allows us to intuitively incorporate prior knowledge on the transition between the models. This also gives us a fine control over the approach. Certainly there is some risk that a user may adjust the values until intended results are obtained. However, in our case this risk is relatively small since the values for the transition probabilities were kept fixed for all our experiments.

### B. Model of Virus Behavior

The behavior of HIV-1 particles throughout the cell entry process is not fully understood and thus this topic is currently the subject of intensive research ([5]). Initially, HIV-1 was assumed to enter the cell exclusively by fusion with the cell membrane. This has been recently challenged by [54], who contend that endocytosis, which is an alternative entry mechanism, also plays a role in the cell entry process. Additionally, the process is dependent on a large number of factors (e.g., sequence of the viral envelope protein, or the presence and density of receptor and co-receptors on the cell surface) whose description would lead to rather intricate models that still need to be established. Here, we restrict ourselves to abstractions of the behavior relative to the inferred temporal intensity models. A straightforward definition for these general behaviors would involve only two abstractions: fusion and non-fusion. We refrain from this approach as it leads to a coarse mapping between the behaviors and the intensity models. While certainly fusion (along with the corresponding temporal intensity models) is the main behavior that we aim to identify, other combinations of temporal intensity models correspond to behaviors that could be of interest to biologists. In our case, we define $N_B = 4$ general behaviors. One behavior is introduced for instances where the intensity corresponding to the outer shell’s label remains constant or increases; we denote this as the Outer Shell Constant or Increasing (OSCI) behavior ($\beta = 0$). The preservation of the label attached to the particle’s inner part is a prerequisite for fusion, and thus we assign another behavior to these types of situations, which are referred as the Inner Part Preservation (IPP) behavior ($\beta = 1$). The loss of the fluorescent label attached to the outer shell is a precondition for fusion too, but also an interesting action in itself, and thus we specify one general behavior to such occurrences; we call this the Outer Shell Loss (OSL) behavior ($\beta = 2$). Finally, both the preservation of the inner part’s label as well as the loss of the outer shell’s label are indicative of fusion (FUSION, $\beta = 3$). The transition probabilities for this 4-state Markov chain could be learned from training data. However, since we are working with high level abstractions of the behavior, we choose manually the transition probabilities for this 4-state Markov chain. This has worked well in practice. For instance, since in our experimental setup the duration of the fusion behavior is relatively short, we have set the probability for the transitions within the fusion behavior to $\phi_{33} = 0.5$. This amounts to an average duration of two frames, which turned out to work well as a prior. The probabilities for transitions from and to the fusion behavior are distributed almost equally. Similar considerations apply to the other behaviors. This yields a flexible model which does not strongly favor transitions towards a particular behavior, where the behaviors have similar durations, and where transitions between all behaviors are allowed, i.e., $\phi_{ij} \neq 0, \forall i, j \in B$. Note that the entries on the diagonal should have higher values than the off-diagonal entries. This reflects the fact that there is usually a higher probability for the particle to remain within the same behavior than to jump to another behavior. Nonetheless the off-diagonal values should be relatively large as this would allow a fast change between behaviors.

To define the measurement model $P(\alpha_t|\beta_t)$ for the HMM, we map the behavior $\beta_t$ as well as the vector $\alpha_t = (\alpha_t^0, \alpha_t^1)$ of the temporal intensity model estimates to a common feature space. The chosen feature space should preserve a certain order for the behaviors, since some behaviors are more closely related (e.g., the FUSION behavior is related to the OSL behavior but it is more distant from the OSCI behavior). To preserve this intuitive ordering over the different behaviors, we map the behaviors to a two-dimensional Hamming space (e.g., [55]), which is a metric space defined by the two-fold Cartesian product of the set $\{0,1\}$, along with the Hamming distance $D_H(\cdot,\cdot)$. The mapping is achieved via the function $h : B \rightarrow \{0,1\}^2$, where $h(\beta)$ is the binary representation of the behavior indices and is defined as follows: $h(0) = (0,0)$, $h(1) = (0,1)$, $h(2) = (1,0)$, and $h(3) = (1,1)$. Similarly, we map the vector $\alpha$ to this Hamming space using the surjective function $g : A^2 \rightarrow \{0,1\}^2$ where $g(\alpha)$ is given in Table I. Note that by adjusting the function $g(\cdot)$ we can cope with different experimental conditions (e.g., in some cases the simultaneous loss of both fluorescent labels may indicate fusion). Our definition of $g(\cdot)$ also implies that channel $c = 0$ corresponds to the intensity of the fluorescent label attached to the outer shell, while channel $c = 1$ represents the intensity of the label tagged to the inner part of the virus particle. To quantify the similarity between $\alpha$ and $\beta$ via $P(\alpha|\beta)$, we use the following exponential function:

$$P(\alpha|\beta) \propto \exp \left( \frac{-D_H(g(\alpha), h(\beta))^2}{2\sigma_H^2} \right),$$

where $\sigma_H$ inversely regulates the level of fidelity in the Hamming space (we used $\sigma_H = 1$).

### IV. Experimental Results

We have evaluated the performance of the layered probabilistic approach using synthetic image sequences as well as...
real microscopy image sequences.

A. Experimental Procedures

To track virus particles in two-channel image sequences, we apply a tracking approach using a Gaussian appearance model as well as model fitting for particle localization, and spatial-temporal filters. An extension to other appearance models (e.g., using the real point-spread function of the microscope) is possible. Note that the image information from both channels is exploited simultaneously by using the measured position estimates computed on both channels. We consider three synthetic scenarios and two real scenarios. For all synthetic image sequences and for all real image sequences of the first scenario, we used independent particle filters for tracking [16]. For the real images of the second scenario, Kalman filters were used (cf. [16]). Besides the positions of single particles, the tracking approach determines the intensity over time $y_{t-1:T}$. To reduce the influence of noise on the measured intensity, the intensity statistic is computed as a weighted mean of the intensities within a spatial neighborhood at the position of the particle, with weights according to a 2D Gaussian function centered at the position of the particle. Note that this approach for computing the intensity statistics is applicable to different localization schemes (spot detection schemes). Since the dimension of our state space is low, $N_s = 1000$ samples for the hybrid particle filter ensure a sufficient support of each posterior model probability. In our experiments, the transition matrix $\Pi$ for the temporal intensity models takes the following form in accordance with the considerations in Section III-A:

$$
\Pi = \begin{pmatrix}
0.8 & 0.1 & 0.1 \\
0.1 & 0.9 & 0.0 \\
0.1 & 0.0 & 0.9
\end{pmatrix}.
$$

Similarly, the transition matrix $\Phi$ for the behaviors follows the considerations in Section III-B and is defined as follows:

$$
\Phi = \begin{pmatrix}
0.5 & 0.2 & 0.2 & 0.1 \\
0.2 & 0.4 & 0.2 & 0.2 \\
0.2 & 0.4 & 0.2 & 0.2 \\
0.1 & 0.2 & 0.2 & 0.5
\end{pmatrix}.
$$

Different values for the variances $Q_{\alpha_t}$ according to the different intensity statistics of each channel are specified. The initial prior probabilities $P(\alpha_0)$ for the three different intensity models are set to $P(\alpha_0) = \begin{pmatrix} 0.7 & 0.15 & 0.15 \end{pmatrix}^T$. Since most particles do not fuse, the prior probabilities $P(\beta_0)$ for the four behaviors are set to $P(\beta_0) = \begin{pmatrix} 0.5 & 0.2 & 0.2 & 0.1 \end{pmatrix}^T$.

For an experimental comparison with a previous approach, we implemented a recent derivative-based approach [23] that exploits the first derivative of the intensity over time. To compute the first derivative with respect to time, finite differences over a local temporal neighborhood are used. Changes in the intensity are detected based on a threshold $T_{\text{deriv}}$ for the absolute value of the calculated derivative values. The sign of the derivative values indicates whether the change is positive or negative. Time steps associated with a positive or negative change are assumed to be governed by a positive intensity change (PIC) or a negative intensity change (NIC) model, respectively. Time steps that are not associated with a change in intensity are assumed to follow a constant intensity (CI) model. The approach determines the underlying behavior of the particle by applying the composite function $(h^{-1} \circ g)(\alpha)$ (cf. Section III-B) that maps the vector $\alpha$ of temporal intensity models computed on both channels to a certain behavior $\beta$.

For each time step $t$, the evaluated approaches compute discrete values (i.e., labels) for the intensity models $\alpha_t = (\alpha^0_t, \alpha^1_t)$ as well as for the behavior $\beta_t$. For each of these three labels, we compute the labeling accuracy $P_{\text{label}}$, which is given as the percentage of correctly labeled time steps relative to all time steps of a single trajectory. The labeling accuracy quantifies the performance over individual time steps. To obtain a measure for individual trajectories, we categorize each trajectory associated with a fusion behavior as a fusion trajectory; otherwise the trajectory is defined as a non-fusion trajectory. Based on these two categories, we compute the accuracy of behavior identification $P_{\text{ident}}$, the identification error $E_{\text{ident}}$, as well as the precision $P_{\text{pre}}$. The first performance measure $P_{\text{ident}}$ reflects the percentage of correctly identified fusion trajectories (True Positives) relative to the number of true fusion trajectories (True Positive Rate or recall/sensitivity). The second measure $E_{\text{ident}}$ is the percentage of false fusion trajectories (False Positives) relative to the number of true non-fusion trajectories (False Positive Rate). The precision $P_{\text{pre}}$ is defined as the ratio between the number of correctly identified fusion trajectories (True Positives) and the number of identified fusion trajectories (True Positives and False Positives). The ground truth consists of a set of trajectories with labels for the intensity models as well as for the behaviors. Each trajectory is also designated as a fusion or a non-fusion trajectory. For the synthetic data, the known intensity statistics of the objects are used to generate the ground truth. For the real data, the ground truth is obtained by manual annotation.

B. Evaluation on Synthetic Images

We evaluate the robustness of our approach using synthetic two-channel image sequences. We compare the performance of our scheme with that of the derivative-based approach [23]. We investigate the following three scenarios. In the first scenario, the aim is to evaluate the performance of the layered approach for the task of detecting fusion at different levels of image noise. This scenario considers single objects. In the second scenario, we model multiple objects and the goal is to examine the capability of the layered approach to retrieve
the number of objects undergoing fusion from a set of objects with heterogeneous behaviors. We also investigate different levels of image noise. In the third scenario, multiple objects are considered and we examine the performance of the layered approach as a function of the performance of the tracking scheme. In all scenarios we render individual objects using a realistic appearance model defined by a 2D Gaussian function:

\[
g(x, y) = I_b + (I_{\text{max}} - I_b) \exp\left(-\frac{(x - x_p)^2 + (y - y_p)^2}{2\sigma_{xy}^2}\right),
\]

where \(I_{\text{max}}\) represents the peak intensity of the object, \(I_b\) corresponds to the background intensity, \((x_p, y_p)\) describe the position of the object, and \(\sigma_{xy}\) regulates the size of the object (we use a value of \(\sigma_{xy} = 1.3\) pixels). We define the signal-to-noise ratio (SNR) as the difference between the peak intensity \(I_{\text{max}}\) of an object and the intensity of the background \(I_b\), divided by the standard deviation of the noise level \(\sigma_n\) ([56]).

The noise model for the intensity is assumed to follow a Poisson distribution. To model different levels of SNR, we set the background intensity to \(I_b = 10\) and vary the peak intensity \(I_{\text{max}}\). In total, we explore seven SNR levels: 11.6, 8.8, 6.5, 4.6, 3.5, 2.8, and 2. In the first and second scenarios, the position \((x_p, y_p)\) over time of individual particles is held constant. In the third scenario, the position over time of individual particles changes and is governed by random walk. The peak intensity \(I_{\text{max}}\) in each channel is varied over time according to the simulated behaviors of the different scenarios.

1) First Synthetic Scenario: In the first synthetic scenario, we examine the performance of the layered approach for identifying the behavior of a single particle undergoing fusion at different SNR levels. For each of the seven SNR levels, we generate 30 two-channel image sequences. Each two-channel image sequence consists of 50 time steps, and each time step includes two images (one for each channel), where each image (16-bit) has dimensions 64×64 pixels. We consider a stationary object positioned at the center of the image. The reason for holding the position constant is to minimize tracking errors, which would otherwise bias the true performance of the layered approach for fusion detection. The peak intensity \(I_{\text{max}}\) of the object in the first channel is modeled using a sigmoid function (mirrored about the y-axis) with a steep transition, which simulates the rapid decrease of the intensity that characterizes the fusion behavior of real virus particles. The peak intensity \(I_{\text{max}}\) in the second channel follows an exponential function. To generate the ground truth labels for the temporal intensity models of each channel, we apply a threshold to the derivative of the underlying intensity model (i.e., the sigmoid or the exponential model). The ground truth labels for the temporal intensity models \(\alpha\) are used for determining the ground truth labels for the behavior \(\beta\) using the composite function \((h^{-1} \circ g)(\alpha)\) (cf. Section III-B). For tracking, we use Gaussian fitting and independent particle filters on each two-channel image sequence. We apply the derivative-based approach as well as the layered approach to the synthetic images. For the derivative-based approach, the threshold values \(T_{\text{deriv}}\) are adjusted according to the SNR level. For the layered approach, the values for the variances \(Q_\alpha\) that regulate the temporal intensity models as well as the noise parameter \(\sigma_n\) are also adjusted based on the SNR level. For both approaches, the parameter values for the first channel are determined empirically so that the approaches detect the negative change entailed by the sigmoid function without inducing too many incorrect labels for the intensity models. The parameter values for the second channel are chosen by taking into consideration that no significant changes in the intensity are expected. For each SNR level, the parameter values are determined based on one image sequence; these values are then used for all 30 image sequences of the corresponding SNR level. The same parameter values are also used for the second and third synthetic scenarios (see Section IV-B2 as well as Section IV-B3 below).

Fig. 1 displays sample images from the two channels at a SNR of 2.8. Note the fast drop in the intensity for the object in the first channel. The original intensities are shown in Fig. 2 together with the true labels for the intensity models of both channels. Although the intensity of the particle drops to the background level in the first channel, the tracking approach determines correctly the position by exploiting the image data from the second channel. The results for the derivative-based scheme, including the calculated derivative values as well as the resulting labels for the intensity models, are shown in Fig. 3. For the first channel, one can see that the decrease in the intensity introduced by the sigmoid model is detected, although the period corresponding to this decrease is rather short. Because of the noise, positive changes in the intensity are also detected. The labeling accuracy is \(P_{\text{label}} = 72\%\) for the first channel while for the second channel the approach achieves a labeling accuracy of \(P_{\text{label}} = 100\%\). Note that in Fig. 3a the original intensity (solid line) represents the peak intensity of a particle while the measured intensity (dashed line) represents a spatial average in the neighborhood of a particle (cf. Section IV-A), therefore the values are typically lower. The results obtained by our layered approach for both channels are displayed in Fig. 4. The intensity for either channel is estimated well. The appropriate intensity models are activated as the intensity fluctuates and the computed sequence of intensity models agrees well with the sequence of true intensity models. Here the labeling accuracy is \(P_{\text{label}} = 86\%\) for the first channel and \(P_{\text{label}} = 100\%\) for the second channel. In comparison to the derivative-based approach, the layered approach yields a less fragmented result for the sequence of intensity models. Also the layered approach provides better results in the presence of noise. In Fig. 5 we show the ground truth labels for the behavior. The labels computed by the derivative-based approach as well as by the layered approach are also shown. Both approaches detect the FUSION behavior, however, the layered approach reconstructs the original labels with a higher fidelity. A certain time lag between the original labels and the labels computed by the layered approach can be observed. This delay is due to the adaptation time of the hybrid particle filter and depends on the transition probabilities in \(\Pi\), the variances \(Q_\alpha\) that regulate the temporal intensity models, as well as the noise parameter \(\sigma_n\). While the adaptation time could be reduced, a compromise between the adaptation time
and the steady behavior of the hybrid particle filter needs to be found. The quantitative values for the labeling accuracy confirm the superior performance of the layered approach, since the derivative-based approach achieves a labeling accuracy of $P_{\text{label}} = 76\%$ while the layered approach attains a labeling accuracy of $P_{\text{label}} = 90\%$.

The labeling accuracy $P_{\text{label}}$ of the evaluated approaches as a function of the SNR is presented in Fig. 6. The diagrams display the mean and standard deviation of the labeling accuracy for the intensity models of the individual channels as well as for the behavior. For each SNR level, the mean and standard deviation are computed over 30 image sequences. For the first channel (see Fig. 6a), the layered approach yields better results than those delivered by the derivative-based approach. For the second channel (see Fig. 6b), the performance for both approaches is equally high. For the behavior (see Fig. 6c), the labeling performance attained by the derivative-based approach is fair. Here, the performance for the behavior is bound by the labeling performance on the intensity models. In general, the degradation of the performance of both approaches relative to the SNR is smooth and monotonic. This indicates that the SNR influences directly the performance of the approaches. In summary, the results show that our layered approach performs well for the task of behavior identification, achieving a fairly good performance at low SNR levels.

Within this scenario we also evaluate the performance of the layered approach as a function of the number of samples $N_s$. We use the same 30 two-channel image sequences as above for SNR = 11.6 and evaluate the result for $N_s = 30, 50, 100, 200, 500$, and 1000. Fig. 7a shows the mean and standard deviation of the labeling accuracy $P_{\text{label}}$ for the first channel as a function of $N_s$. It can be seen that the performance of the approach is fairly good even for a relatively low number of samples (e.g., $N_s = 50$). The reason for this is that the dimension of the hybrid state space is relatively low. However, as indicated by the error bars, the standard deviation is relatively large for a low number of samples. Thus a larger number of samples (e.g., $N_s \geq 500$) is required to obtain robust results (in all our experiments we used $N_s = 1000$). A similar trend is observed for the labeling accuracy $P_{\text{label}}$ for the second channel as well as for the behavior (see Figs. 7b, 7c). Fig. 8 shows how the error $(1 - P_{\text{label}})$ decreases with an increasing number of samples $N_s$. Here $N_s$ takes values between 10 and 200 with an increment of 10.

2) Second Synthetic Scenario: In the second scenario, we consider multiple objects and the goal is to study the performance of the layered approach for determining fusion given a set of particles with different behaviors. In this scenario, we generate two-channel image sequences consisting of 50 time steps. The images $(256 \times 256$ pixels, 16-bit) display 30 objects rendered with a 2D Gaussian appearance model (see (27)). The image positions of the objects are randomly chosen and their positions remain constant over time. For each SNR level we generate 30 two-channel image sequences. The peak intensity $I_{\text{max}}$ of each object in the first channel is modeled by either a constant, linear, exponential, or sigmoid function. In real microscopy images, photobleaching induces a slight decrease in the intensity. To simulate this phenomenon in our experiment, the linear and exponential models describe a slight decrease in the intensity over time as well. Except for the sigmoid function, similar models are used for the peak intensity $I_{\text{max}}$ in the second channel, and thus no sharp decreases in the intensity are visible in this channel. The models are chosen randomly for each object. If for a certain object the sigmoid model is selected, the corresponding object is assumed to undergo fusion. In this scenario, 12 out of the 30 objects exhibit a fusion behavior. Fig. 9 displays sample

![Fig. 1](image1.png)

First synthetic scenario: Sample time steps of a synthetic image sequence. The SNR is 2.8. The intensity over time of the particle in the first channel is governed by a sigmoid model. The intensity over time in the second channel is given by an exponential model. For visualization purposes, the image contrast has been enhanced and the image intensities have been inverted.

![Fig. 2](image2.png)

First synthetic scenario: Ground truth for the intensity as well as for the intensity models.
image sections from both channels at a SNR of 2.8. One can observe a significant reduction in the intensity of the objects. We apply the tracking approach to each two-channel image sequence. We use the same parameter values from the first synthetic scenario for the derivative-based approach approach as well the for layered approach.

For each image sequence, we compute the mean labeling accuracy $P_{\text{label}}$, which is defined as the mean of the labeling accuracy $P_{\text{label}}$ obtained over all 30 objects. Then we compute the mean values and standard deviations for $P_{\text{label}}$ over all 30 image sequences for the temporal intensity models in both channels as well as for the behavior. These values are presented in Fig. 10. Overall for both approaches the labeling performance for the intensity models as well as for the behavior is above 80% for all SNR levels. For example, at a SNR of 2.8 the derivative-based approach achieves a mean value for $P_{\text{label}}$ of approximately 86% for the behavior while the layered approach attains a mean value for $P_{\text{label}}$ of 90%. The layered approach outperforms generally the derivative-based approach. The performance degrades smoothly for both approaches as the SNR decreases. In this scenario we also compute the accuracy of behavior identification $P_{\text{ident}}$, the identification error $E_{\text{ident}}$, as well as the precision $P_{\text{pre}}$ as a function of the SNR. While $P_{\text{label}}$ is relatively tolerant to errors in the computed labels for the behavior, $P_{\text{ident}}$, $E_{\text{ident}}$, as well as $P_{\text{pre}}$ are relatively sensitive to such errors. As such, a high labeling accuracy $P_{\text{label}}$ for the behavior may not necessarily correlate with a high accuracy of behavior identification $P_{\text{ident}}$, a high precision $P_{\text{pre}}$, or conversely with a low identification error $E_{\text{ident}}$. The mean and standard deviation of these measures computed over 30 image sequences per SNR level are shown in Fig. 11. The derivative-based approach performs rather poorly in terms of $P_{\text{ident}}$. As the SNR decreases, the identification accuracy decreases, i.e., the ability of the approach to identify particles undergoing fusion is hampered by the noise. In terms of the identification error $E_{\text{ident}}$, the approach yields a large number of false positives, even at high SNR levels. In other words, a large percentage of non-fusion trajectories are mistakenly identified as fusion trajectories. The low precision $P_{\text{pre}}$ also indicates that the number of correctly identified fusion trajectories relative to the number of false positives is low, especially at low SNR levels. The layered approach provides a better performance. The accuracy of behavior identification $P_{\text{ident}}$ is close to 100% down to a SNR of approximately 3.5. Likewise, the identification error $E_{\text{ident}}$ remains below 22% down to a SNR of approximately 3.5. Below a SNR of 3.5 the accuracy of behavior identification $P_{\text{ident}}$ decreases rather sharply. Nonetheless, the identification error $E_{\text{ident}}$ remains relatively small, which is a favorable property because, while not all true fusion trajectories might be retrieved, the approach also delivers fewer false positives. The precision $P_{\text{pre}}$ shows that at large SNR levels the approach identifies fusion trajectories without incurring a relatively large number of false positives. At lower SNR levels the precision $P_{\text{pre}}$ decreases relatively strongly due to the increase in the number of false positives. Note that in Fig. 11a, at low SNR levels, the derivative-based approach achieves a higher $P_{\text{ident}}$ compared to the layered approach. This is because at low SNR levels the derivative-based approach (due to its sensitivity to noise) considers most
trajectories as fusion trajectories, i.e., the approach correctly identifies most true fusion trajectories (high $P_{\text{ident}}$), but it also considers a large number of non-fusion trajectories as fusion trajectories (high $E_{\text{ident}}$), which is unfavorable. The poorer performance of the derivative-based approach is also reflected by the significantly lower precision $P_{\text{pre}}$ compared to the layered approach. In general, the results suggest that the layered approach identifies fusion fairly well given a set of objects exhibiting different behaviors.

3) Third Synthetic Scenario: We also evaluate the performance of the layered approach as a function of the performance of the tracking approach. We generate two-channel image sequences consisting of 50 time steps. The images have dimensions $256 \times 256$ pixels (16-bit). To control the performance of the tracking approach, we vary the number of objects in the images using $n_{\text{obj}} = 30, 50, 100, 150, 200, 250, 300, 400,$ and $500$. The higher $n_{\text{obj}}$, the higher the object density and thus more tracking errors occur. The initial image position of each object is random and the motion is governed by random walk. The peak intensity $I_{\text{max}}$ of each object in each channel is modeled as in the second synthetic scenario (cf. Section IV-B2) and we use $\text{SNR} = 11.6$. For each number of objects $n_{\text{obj}}$, we generate 30 two-channel image sequences (thus in total we use 270 image sequences). We apply a tracking approach based on independent particle filters [16] to each two-channel image sequence. To quantify the tracking performance, we calculate the linking error $E_{\text{link}}$ [57], [58].
which is defined as:

$$E_{\text{link}} = 1 - \frac{n_{\text{links,correct}}}{n_{\text{links,total}}}$$  \hfill (28)

where $n_{\text{links,correct}}$ is the number of correct links and $n_{\text{links,total}}$ is the number of true links. A link corresponds to a displacement vector between two consecutive positions of a trajectory. A correct link corresponds to a displacement vector between two positions close to two consecutive positions of a true trajectory. We apply the layered approach for behavior identification using the computed trajectories. We use the same parameter values for the layered approach as in the first and second synthetic scenarios.

For each number of objects $n_{\text{obj}}$, we compute the mean value for $E_{\text{link}}$ over the 30 image sequences (see Fig. 12). It can be seen that $E_{\text{link}}$ increases linearly with $n_{\text{obj}}$ and that the tracking approach copes relatively well with a large number of objects (e.g., for $n_{\text{obj}} = 500$, $E_{\text{link}} = 18\%$).

We also determine the accuracy of behavior identification $P_{\text{ident}}$, the identification error $E_{\text{ident}}$, as well as the precision $P_{\text{pre}}$ as a function of $E_{\text{link}}$. The result in Fig. 13a shows that $P_{\text{ident}}$ degrades slowly as $E_{\text{link}}$ increases. In Fig. 13b, $E_{\text{ident}}$ exhibits a linear relation with $E_{\text{link}}$ which suggests that the number of false positives is directly influenced by tracking errors. The precision $P_{\text{pre}}$ (Fig. 13c) decreases as $E_{\text{link}}$ increases. To summarize, the layered approach copes well with a small number of tracking errors. For a larger
Fig. 9. Second synthetic scenario: Sample time steps of a synthetic image sequence. The SNR is 2.8. For visualization purposes, the image contrast has been enhanced and the image intensities have been inverted.

Fig. 10. Second synthetic scenario: Mean labeling accuracy $P_{\text{label}}$ as a function of the SNR for the intensity models (first and second channel) as well as for the behavior. The mean values (and standard deviations) of a derivative-based approach (‘Derivative’) as well as the mean values (and standard deviations) of the layered probabilistic approach (‘Layered’) are shown.

Fig. 11. Second synthetic scenario: Performance in terms of the identification accuracy $P_{\text{ident}}$, identification error $E_{\text{ident}}$, and the precision $P_{\text{pre}}$ as a function of the SNR. The mean values (and standard deviations) of a derivative-based approach (‘Derivative’) as well as the mean values (and standard deviations) of the layered probabilistic approach (‘Layered’) are shown.

Fig. 12. Third synthetic scenario: Performance of the tracking approach in terms of the linking error $E_{\text{link}}$ as a function of the number of objects $n_{\text{obj}}$. The mean values (and standard deviations) of a tracking approach based on independent particle filters (‘Particle Filter’) [16] are shown.
number of tracking errors, the performance degrades, however, the degradation is smooth.

C. Evaluation on Real Microscopy Images

We have also applied the layered approach to real microscopy image sequences displaying HIV-1 particles. We have considered two scenarios each involving a different technique for fluorescently labeling the virus particles yielding different kinds of image data. We have carried out an experimental comparison with a derivative-based approach [23]. In addition, we perform a comparison with a trajectory-based approach (cf. [16]) that compares the endpoints of trajectories obtained by tracking using a single channel with the endpoints of trajectories obtained by tracking using two channels. A fusion event is identified when the endpoint of a single-channel trajectory differs from the endpoint of a corresponding two-channel trajectory. Since this approach based on trajectory endpoints only takes into consideration spatial information, it cannot determine the occurring temporal intensity models of an individual particle. Thus $P_{\text{label}}$ for the intensity models cannot be calculated. Likewise, since the behaviors are defined in terms of the intensity models (see Section III-B), the occurring behavior for each time step could not be determined and so $P_{\text{label}}$ for the behavior is not computed. Since the approach based on trajectory endpoints categorizes each trajectory as a fusion or non-fusion trajectory, we report results for $P_{\text{ident}}$, $E_{\text{ident}}$, and $P_{\text{pre}}$.

1) First Real Scenario: In the first real scenario, the outer shell (viral matrix) of individual pseudotyped HIV-1 particles is labeled with an enhanced green fluorescent protein (MA.eGFP) while the inner core part (viral protein R) is tagged with the red fluorescent protein (mRFP.Vpr). The HIV-1 particles are pseudotyped with the glycoprotein of the vesicular stomatitis virus (VSV-G) and incubated with HeLa cells [7]. Images are acquired using a Zeiss Axiovert 200 M microscope with a Roper Scientific Cascade II EM-CCD. A pair of images (one per channel) is recorded every 100 ms. Upon fusion, the label attached to the outer shell dissolves, and thus a decrease in the intensity in the corresponding channel is observed. The acquired image sequences consist of 200-400 two-channel images (512×512 pixels; 16-bit). Within this scenario, we evaluate 5 image sequences. To track the virus particles in the two-channel image sequences, we use Gaussian fitting and independent particle filters [16]. Ground truth for the intensity labels is obtained manually for 20 of the computed trajectories within each image sequence by inspecting the intensity over time of the tracked virus particles. The labels for the behavior are computed via the composite function $(h^{-1} \circ g)(\alpha)$ on the manually determined intensity models $\alpha$. Based on the labels for the behavior, trajectories are categorized as fusion or non-fusion trajectories; see Table II. For the evaluation, the parameter values for the derivative-based approach as well as for the layered approach are kept fixed for all image sequences.

Fig. 14. First real scenario: Tracking results for the real image sequence “Sequence 1”. The time step is $t = 199$. For visualization purposes, the image contrast has been enhanced and the image intensities have been inverted.

An image from Sequence 1 in Table II and the correspond-
ing tracking results are shown in Fig. 14. A particle undergoing fusion is displayed in Fig. 15. The ground truth labels for the intensity models for either channel reflect a decrease in the intensity observed in the first channel that corresponds to fusion (see Fig. 16a) as well as an increase in the intensity in the second channel (see Fig. 16b). The corresponding changes in the intensity are illustrated in Fig. 17a and Fig. 17d, where the measured intensity over time computed for each channel is shown. The computed derivative values for the first channel as well as for the second channel are shown in Fig. 17b and Fig. 17e, respectively. The results for the derivative-based approach for the estimated intensity models for the first and second channels are shown in Fig. 17c and Fig. 17f. Qualitatively, the derivative-based approach does not recover very well the motif underlying the true sequence of the intensity models. In particular, wrong labels scattered throughout the computed sequences of the intensity models lead to fragmented segments within these sequences. Relative to the ground truth labels, the derivative-based approach achieves a labeling accuracy of $P_{\text{label}} = 92\%$ for the intensity models of the first channel and $P_{\text{label}} = 94\%$ for the intensity models of the second channel. In comparison, the layered approach retrieves the underlying motif relatively well (cf. Fig. 18c and Fig. 18f). Note the timely activation of the appropriate intensity models (cf. Fig. 18b and Fig. 18e). Accordingly, the approach achieves a labeling accuracy of $P_{\text{label}} = 96\%$ for the first channel and $P_{\text{label}} = 95\%$ for the second channel. Fig. 19 displays the ground truth labels for the behavior as well as the labels for the behavior computed by the derivative-based approach as well as by the layered approach. Both approaches identify the fusion behavior. For the derivative-based approach, the fragmentation observed in the sequences of intensity models carries over to the sequence of behaviors. This leads to a lower labeling accuracy for the behavior ($P_{\text{label}} = 87\%$). In contrast, the layered approach yields fairly accurate results, as reflected by a higher labeling accuracy for the behavior ($P_{\text{label}} = 95\%$).

Table III displays the performance of the derivative-based scheme as well as the layered approach in terms of the mean labeling accuracy $P_{\text{label}}$ for the intensity models on both channels as well as for the behavior. The mean and standard deviation over all five two-channel image sequences are shown as well. Overall, the labeling accuracy for the intensity models is quite high for both approaches (the mean value for $P_{\text{label}}$ is above $90\%$). For the behavior, the derivative-based scheme yields a mean value for $P_{\text{label}}$ of $88\%$ (standard deviation of $5\%$) while the layered approach delivers a mean value for $P_{\text{label}}$ of $93\%$ (standard deviation of $5\%$). In all sequences, the layered approach achieves a better labeling performance than the derivative-based approach. The accuracy of behavior identification $P_{\text{ident}}$, the identification error $E_{\text{ident}}$, as well as the precision $P_{\text{pre}}$ over all sequences are shown in Table IV. For instance, for Sequence 1 we obtain $P_{\text{ident}} = 100\%$ for the derivative-based approach, for the approach based on trajectory endpoints, as well as for the layered approach.

The derivative-based approach yields an identification error of $E_{\text{ident}} = 58\%$ while the approach based on trajectory endpoints incurs $E_{\text{ident}} = 5\%$. The reason for the high value of $E_{\text{ident}}$ for the derivative-based approach is that this approach is quite sensitive to image noise. The approach based on trajectory endpoints in comparison does not directly evaluate the image intensities. However, this approach also delivers a certain number of false positives. In comparison, the layered approach achieves $E_{\text{ident}} = 0\%$. The precision of the derivative-based approach ($P_{\text{pre}} = 8\%$) also reflects the larger number of false positives generated by this approach. The approach based on trajectory endpoints yields $P_{\text{pre}} = 50\%$. A higher precision is achieved by the layered approach ($P_{\text{pre}} = 100\%$). The other image sequences studied within this scenario do not display particles undergoing fusion and therefore $P_{\text{ident}}$ is not calculated. For the identification error $E_{\text{ident}}$, the derivative-based approach yields a mean value of $E_{\text{ident}} = 37\%$ (standard deviation of $13\%$). The values for $E_{\text{ident}}$ of the derivative-based approach are thus in accordance with the results obtained for the second synthetic scenario (cf. Fig. 11), where the derivative-based approach yielded a large identification error $E_{\text{ident}}$. In comparison, the approach based on trajectory endpoints achieves a mean value of $E_{\text{ident}} = 25\%$ (standard deviation of $15\%$). This suggests that this approach yields a lower number of false positives compared to the derivative-based approach. The layered approach achieves the lowest mean value for the identification error ($E_{\text{ident}} = 0\%$, standard deviation of $0\%$) and the result is in agreement with the results obtained for the synthetic images (cf. Fig. 11), where $E_{\text{ident}}$ of the layered approach is significantly lower compared to the derivative-based approach. Because some image sequences do not display fusion particles, and because in those cases the derivative-based approach as well as the

<table>
<thead>
<tr>
<th>Sequence</th>
<th>1st Ch.</th>
<th>2nd Ch.</th>
<th>Behavior</th>
<th>1st Ch.</th>
<th>2nd Ch.</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96</td>
<td>94</td>
<td>87</td>
<td>98</td>
<td>97</td>
<td>93</td>
</tr>
<tr>
<td>2</td>
<td>99</td>
<td>96</td>
<td>95</td>
<td>100</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>3</td>
<td>99</td>
<td>89</td>
<td>85</td>
<td>100</td>
<td>97</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>94</td>
<td>88</td>
<td>82</td>
<td>96</td>
<td>91</td>
<td>86</td>
</tr>
<tr>
<td>5</td>
<td>97</td>
<td>92</td>
<td>91</td>
<td>98</td>
<td>95</td>
<td>95</td>
</tr>
</tbody>
</table>

Mean: 97 | 92 | 88 | 98 | 96 | 93
Std. Dev.: 2 | 3 | 5 | 1 | 3 | 5

TABLE III

First real scenario: Results for real image sequences in terms of the mean labeling accuracy $P_{\text{label}}$ [%] for the temporal intensity models in both channels as well as for the behavior. The mean values and standard deviations are also shown.
First channel, $t = 0$
First channel, $t = 40$
First channel, $t = 80$
Second channel, $t = 0$
Second channel, $t = 40$
Second channel, $t = 80$

Fig. 15. First real scenario: Tracking results for a virus particle undergoing fusion. For visualization purposes, the image contrast has been enhanced and the image intensities have been inverted.

Fig. 16. First real scenario: Ground truth for the intensity models.

(a) True labels for the intensity models (first channel)
(b) True labels for the intensity models (second channel)

2) Second Real Scenario: In the second real scenario, the outer shell (lipid envelope) of HIV-1 particles pseudotyped with the Env glycoprotein of the ecotropic murine leukemia virus is labeled with the yellow fluorescent protein (ML-VEnv·YFP). The mCherry fluorescent protein (MA.mCherry) is attached to the inner part (viral matrix) [8]. Images are acquired using the same wide-field microscopy setup (but in TIRF mode) as in the first real scenario. Here, the acquisition rate for a pair of images is 2 seconds, which experimentally was found to be sufficient. Fusion entails a loss of the label attached to the outer shell (lipid envelope) along with a decrease in intensity in the corresponding channel. However, a decrease in intensity in both channels is also indicative of fusion when using this labeling strategy. This variant is accommodated into the evaluated approaches by adjusting the function $g(\alpha)$. We evaluate 8 image sequences. We use an approach based on Kalman filters to track the virus particles (cf. [16]). The ground truth for the fusion events is determined.

### Table IV

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Derivative</th>
<th>Trajectory Endpoints</th>
<th>Layered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_{\text{ident}}$</td>
<td>$E_{\text{ident}}$</td>
<td>$P_{\text{pre}}$</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>58</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>–</td>
<td>35</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>–</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>–</td>
<td>35</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>–</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>–</td>
<td>37</td>
<td>2</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>–</td>
<td>13</td>
<td>4</td>
</tr>
</tbody>
</table>

### Table V

<table>
<thead>
<tr>
<th>Sequence</th>
<th>No. of evaluated trajectories</th>
<th>No. of fusion trajectories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>1</td>
</tr>
</tbody>
</table>
Fig. 17. First real scenario: Results obtained using a derivative-based approach. The labeling accuracy is $P_{\text{label}} = 92\%$ for the first channel and $P_{\text{label}} = 94\%$ for the second channel.

Fig. 20. Second real scenario: Tracking results for the real image sequence “Sequence 2”. The time step is $t = 99$. For visualization purposes, the image contrast has been enhanced and the image intensities have been inverted.

As fusion occurs, the intensity drops almost simultaneously in both channels. Accordingly, the ground truth labels for the intensity models for both channels include concurrent time periods where the ‘Negative Intensity Change’ (NIC) model is dominant (see Fig. 22). The measured intensity for both channels shown in Fig. 23a and Fig. 23d reflects more explicitly the aforementioned drop in intensity. Here the computed derivative values are very noisy (Fig. 23b and Fig. 23e) and consequently the derivative-based approach yields fragmented sequences for the estimated intensity models (Fig. 23c and Fig. 23f). The labeling accuracy for the intensity models is $P_{\text{label}} = 75\%$ for the first channel and $P_{\text{label}} = 91\%$ for the second channel. The results obtained by the layered approach are displayed in Fig. 24. The approach achieves $P_{\text{label}} = 88\%$ for the first channel and $P_{\text{label}} = 85\%$ for the second channel. Thus, the result for the first channel is much better. The reason for the lower performance for the second channel is that the hybrid particle filter did not adapt quickly enough to the fusion-induced change. This could be addressed by choosing different values for the parameters (e.g., the noise parameter $\sigma_n$). However, the adaptation time and the steady behavior of the hybrid particle filter need to be balanced. The ground truth labels for the behavior of this particle are shown in Fig. 25. For the derivative-based approach erroneous labels for the intensity models computed on either channel are propagated to the labels for the behavior. Accordingly, the labeling performance for the behavior is $P_{\text{label}} = 74\%$. For the layered approach, the computed behaviors are in better agreement with the ground truth labels yielding a significantly higher labeling accuracy of $P_{\text{label}} = 87\%$. 

manually. A summary of the real image data is shown in Table V.

As an example, an image from Sequence 2 in Table V and the respective tracking results are shown in Fig. 20. Notwithstanding the relatively high object density, the tracking approach determines successfully the trajectories of the virus particles. A virus particle undergoing fusion is shown in Fig. 21. In this case, the contrast of the first channel is much lower, which entails subtler changes in the intensity.
Fig. 18. First real scenario: Results obtained using a hybrid particle filter. The labeling accuracy is $P_{\text{label}} = 96\%$ for the first channel and $P_{\text{label}} = 95\%$ for the second channel.

Fig. 19. First real scenario: Ground truth for the behavior and results obtained using a derivative-based approach as well as the layered probabilistic approach. The labeling accuracy for the derivative-based approach is $P_{\text{label}} = 87\%$ while for the layered approach the labeling accuracy is $P_{\text{label}} = 95\%$.

As shown in Table VI, the overall labeling accuracy for the derivative-based approach in this scenario is worse compared to the first real scenario (see Table III). The reason for this is that in the second real scenario the changes in the intensity corresponding to fusion are smaller. To detect such weaker changes, the threshold applied to the derivative values has to be lowered. Because of the noise, the low threshold leads to a larger number of errors for the labels of the intensity models. The errors in the labels for the intensity models are propagated to the labels for the behavior, and thus the performance of this approach suffers. Concretely, the mean value of $P_{\text{label}}$ for the intensity models as well as for the behaviors is below 90%. The layered approach instead reconstructs the underlying intensity via the proposed intensity models. A good reconstruction enhances the dynamic properties of the underlying signal, which leads to a correct identification of the dominant intensity models. Correct identification of the intensity models in turn provides improved predictions for the intensity at subsequent time steps. This indicates that the estimates for the intensity and the intensity models benefit from the joint estimation process embodied by the stochastic hybrid system. For the layered approach, the mean value of $P_{\text{label}}$ for the intensity models as well as for the behaviors is above 95%. In Table VII, the performance of the approaches (and that of the approach based on trajectory endpoints) is further reflected by the accuracy of behavior identification $P_{\text{ident}}$, the identification error $E_{\text{ident}}$, as well as the precision $P_{\text{pre}}$ computed for all eight image sequences. Here, the derivative-based approach as well as the layered approach achieve a mean accuracy of behavior identification of $P_{\text{ident}} = 94\%$ (standard deviation of
18% for both approaches). That is, both approaches recover well the true fusion trajectories. In comparison, the approach based on trajectory endpoints yields $P_{ident} = 62\%$ (standard deviation of 44%). The reason for this is that this approach assumes that fusion is described by a decrease in intensity in a single channel only. In this scenario, however, fusion is described by a decrease in intensity in both channels.

For the identification error $E_{ident}$, the derivative-based approach delivers a mean value of $E_{ident} = 75\%$ (standard deviation of 36%), which further confirms the sensitivity of the approach to the image noise. The result for the approach based on trajectory endpoints is moderate ($E_{ident} = 18\%$, standard deviation of 14%). The best result is achieved by the layered approach ($E_{ident} = 5\%$, standard deviation of 7%).

In terms of the precision $P_{pre}$, the derivative-based approach achieves a mean value of $P_{pre} = 12\%$ (standard deviation of 10%), where the low precision may be attributed to the large number of false positives. The approach based on trajectory endpoints also yields a low precision ($P_{pre} = 18\%$, standard deviation of 11%), which can be explained by the low number of correctly identified fusion trajectories. The layered approach achieves $P_{pre} = 70\%$ (standard deviation of 33%), which highlights the improved performance compared to previous approaches. Overall, the results suggest that the layered approach is well suited for the task of identifying fusion of HIV-1 particles.

The approach was implemented in Java within our software ViroTracker ([16]). The computation time of the approach scales linearly with respect to the number of samples for the hybrid particle filter and quadratically with respect to the number of states in the top-most Markov chain (i.e., number of behaviors). For example, the computation time for one trajectory comprising 100 time steps and using 1000 samples is ca. 1 second on an AMD Opteron (2.3 GHz) CPU running...
Fig. 23. Second real scenario: Results obtained using a derivative-based approach. The labeling accuracy is $P_{\text{label}} = 75\%$ for the first channel and $P_{\text{label}} = 91\%$ for the second channel.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Derivative</th>
<th>Trajectory Endpoints</th>
<th>Layered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_{\text{ident}}$</td>
<td>$E_{\text{ident}}$</td>
<td>$P_{\text{pre}}$</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>83</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>26</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>11</td>
<td>33</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>83</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>Mean</td>
<td>94</td>
<td>75</td>
<td>12</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>18</td>
<td>36</td>
<td>10</td>
</tr>
</tbody>
</table>

TABLE VII
SECOND REAL SCENARIO: RESULTS FOR REAL IMAGE SEQUENCES OBTAINED IN TERMS OF THE ACCURACY FOR BEHAVIOR IDENTIFICATION $P_{\text{ident}}$ [%], THE IDENTIFICATION ERROR $E_{\text{ident}}$ [%], AND THE PRECISION $P_{\text{pre}}$ [%]. THE MEAN VALUES AND STANDARD DEVIATIONS ARE SHOWN, TOO.

Linux.

V. CONCLUSIONS

We introduced a new approach for the identification of behaviors of interest for single virus particles in two-channel fluorescence microscopy image sequences. Our approach is based on a layered architecture formulated within a Bayesian estimation framework. We use a stochastic hybrid system for modeling the intensity and the temporal intensity models. Our approach also employs a hidden Markov model (HMM) to represent the top-level behaviors of a single particle. The stochastic hybrid system and the HMM are combined using a maxbelief approach. Inference is carried out with a hybrid particle filter as well as with the Viterbi algorithm. In comparison to previous approaches for identifying behaviors of fluorescent particles, the approach does not assume a constant behavior for the particles, and thus the approach can cope with objects exhibiting heterogeneous behaviors. Also, we take into account the temporal coherence of the particle’s behavior via the transition probabilities defined for the Markov chains underlying the stochastic hybrid system and the HMM. By adopting a probabilistic framework, we account for the inherent uncertainty involved in the behavior of single virus particles. The presented approach is mainly parametrized by the variances $Q_\alpha$, for the temporal intensity models of each channel, the noise parameter $\sigma_n$, as well as the number of samples $N_s$. The variances $Q_\alpha$ as well as the noise parameter $\sigma_n$ could be determined based on the intensity values of sample trajectories via automatic parameter estimation schemes. As suggested by our experimental results,
the number of samples $N_s$ should be relatively large (e.g., $N_s \geq 500$) to obtain robust results. We have applied our layered approach to synthetic image sequences as well as to real image sequences. We also have compared the performance of our approach with that of a previous derivative-based approach. In addition, we have performed a comparison with an approach based on trajectory endpoints. The results on two-channel synthetic images demonstrated the applicability of our layered approach for identifying the intensity models as well as the behavior of single particles. By exploring a range of different levels of noise, we characterized the performance of our approach as a function of the SNR level. The results established that our approach performed well at typical SNR levels and relatively well at low SNR levels. Also the performance of the approach degraded gracefully as the SNR decreased. In addition, we evaluated the performance of the approach relative to the performance of the tracking approach. It turned out that the layered approach tolerates well minor tracking errors and that the performance degrades smoothly with larger tracking errors. The application of the layered approach to real microscopy image data showed that the approach is well suited for detecting fusion of HIV-1 particles with the cell membrane. Overall, the layered approach yielded a significant improvement in the performance compared to a derivative-based approach, which relies entirely on data-driven information. Instead, our approach integrates model-driven information (e.g., predictions based on the temporal intensity models) with the observed intensity information. The layered approach also outperformed the approach using trajectory endpoints. This highlights the benefits of directly analyzing the
image intensities of individual particles. Certainly, the layered approach has also limitations. Since a compromise between the adaptation time and the steady behavior of the hybrid particle filter has to be found, the labels computed by the approach may exhibit a certain time delay. This time delay can be reduced by adjusting the transition probabilities in the matrix $\Pi$, the variance parameters $Q_\alpha$ as well as the noise parameter $\sigma_\alpha$. However, doing this for single image sequences might be impractical. Alternatively, these parameters could be optimized based on the image data. We note that there is some risk that a user may adjust the transition probabilities until intended results are obtained. To reduce this risk, the transition probabilities should be kept fixed for all experimental results as done in our case. Also, there is some risk of overfitting. However, since in our case the models are not too complex (e.g., each intensity model is parametrized by only a single parameter) this risk is low. Another issue is to improve the performance of the layered approach in case of severe tracking errors. Current and future work includes the interpretation of the results delivered by the layered approach in terms of their biological significance. Also in future work, the layered approach could be extended for other applications using image sequences with more than two channels.

**References**


