EUSIPCO 2013 1569746181

A MAXIMUM LIKELIHOOD METHOD FOR LIFETIME ESTIMATION IN PHOTON COUNTING-BASED FLUORESCENCE LIFETIME IMAGING MICROSCOPY

A. Chessel, F. Waharte, J. Salamero

UMR 144 CNRS - PICT IBiSA
26 rue d'Ulm
75 248 Paris cedex 05, France

C. Kervrann

Inria, Centre de Rennes - Bretagne Atlantique
Campus Universitaire de Beaulieu
35 042 Rennes Cedex, France

ABSTRACT

In this paper we derive a Maximum Likelihood (ML) framework for photon counting-based fluorescence lifetime estimation in Fluorescence Lifetime Imaging Microscopy (FLIM) from the biophysical phenomenon and instrument models. Data collected at a given pixel consist of photon counts exponentially decreasing along time and are assumed to follow Poisson statistics. Both pointwise approaches and a neighborhood-wise approach are proposed to take explicitly into account the spatial correlation of data. Evaluations and comparisons are presented on simulated as well as on experimental biological image data.

Index Terms— Photon counting, Poisson statistics, liftetime estimation, fluorescence microscopy

1. INTRODUCTION

Fluorescence microscopy and Green Fluorescent Protein (GFP) tagging have become widely used tools of modern cellular biology. Unlike traditional fluorescence imaging techniques, FLIM imaging aims at estimating fluorescence lifetime [3], that is the average time during which the fluorescent molecule stays in the excited state before returning to the ground state by emitting a photon [7]. Typical biological applications include imaging spectrally indistinguishable fluorescing species or measuring molecular proximity between two fluorophores by Förster Resonance Energy Transfer (FRET). Fluorescence lifetime measurements were first exploited in spectroscopy on homogeneous samples with a high precision to characterize the photophysical properties of a specific compound. It is now routinely used in biology, but mainly to analyse images from dimmer and heterogeneous fluorescent samples.

In FLIM imaging, at each pixel the response of the fluorescent sample to a periodic excitation light is measured (repetition of pulses for time-domain approaches or modulated light for frequency domain techniques). In the time-domain method, this response is the number of detected photons decreasing in time (see Fig. 1). This decay in fluorescence can be described by a first-order reaction and thus can be modeled by an exponential with a fixed rate, called the *fluorescence lifetime* of the fluorophore. In our study, we focus on the Time-Correlated Single Photon Counting (TCSPC) technique [6] so the measurement corresponds at each time point to the number of photons reaching the detector and obeys Poisson statistics.

The most standard approach to estimate lifetime is based on a least-square fitting of a mono-exponential (or multi-exponential) function to the real data [8] and the Poisson distribution is generally assumed to be approximately Gaussian. Nevertheless, the Maximum Likelihood approach is probably more recommended as for many image reconstruction problems. In photon counting-based image reconstruction, ML is known to be equivalent to minimizing the Kullback-Leibler (KL) divergence of the computed data from the acquired data. Nevertheless, taking into account the spatial coherence the data and image smoothness as considered in our approach, is especially relevant in the case of very low photon counts for applications in cell biology.

The remainder of the paper is organized as follows. In Section 2, we present the theory and models in TCSPC-FLIM imaging. In Section 3, we formulate the lifetime estimation problem as a pointwise and neighborhood-wise ML problem. In Section 4 the performance of the algorithm is evaluated on synthetic and experimental data. The proposed method allows for recovering the spatial lifetime information and enables to demonstrate experimentally localized spatial interaction of the two proteins of interest on endosomes.

2. TCSPC-FLIM IMAGING

In this section, we present the theory in TCSPC FLIM imaging, now routinely used in biology for lifetime estimation and spatially-resolved estimation of nanometer-scale molecular proximity between proteins.

Biophysical models of fluorescence A fluorescence substance can absorb a photon of a given wavelength and remit a very short time later (with a picosecond (ps) to a nanosecond (ns) delay) a photon at another wavelength (corresponding to

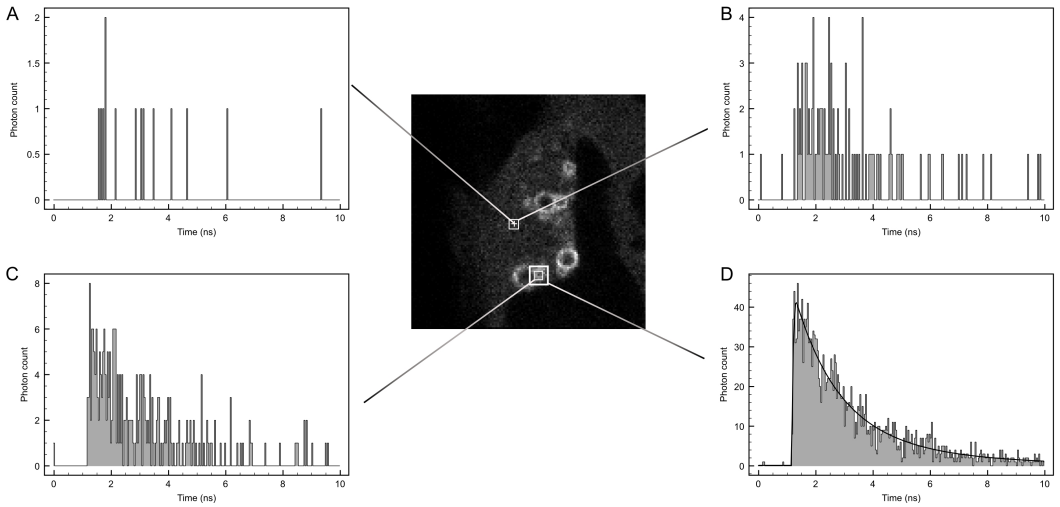

Fig. 1. Example of typical TCSPC FLIM data. Total fluorescence intensity is shown in the center and corresponds to the sum of photon counts along the time axis at each pixel. The four side graphs correspond to time dependent photon counts in four different regions with variable sizes. By considering large regions, we observe an exponential fluorescence decay (see D). A: one pixel region; B and C: 3×3 patches at different locations; D: 15×15 patch and liftetime estimation by least-square fitting (commercial software).

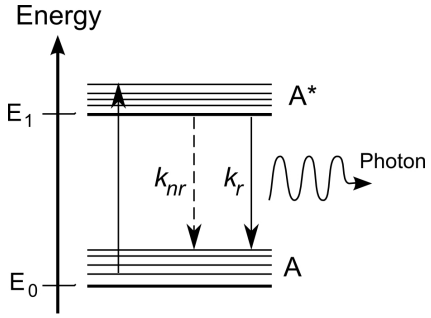

Fig. 2. Simplified Perrin-Jablonski diagram of fluorescence. From a fundamental state A, the fluorophore is excited to a state A^* by absorbing a photon at a suitable wavelength. The return to fundamental state can be either radiative, with a photon emission or non-radiative. The fluorescence lifetime is the average duration of the state A^*.

a lower energy). The Perrin-Jablonski diagram shown in Fig. 2, is considered for energetic interpretation: when absorbing a photon, the compound A goes from the ground state to an excited state A^*. After loosing some energy by vibrational relaxation, it goes down to the fundamental state. It can do so by radiative de-excitation, i.e. emitting a photon (the fluorescence phenomenon itself), or by other non-radiative means.

To those two processes are associated the two reaction rates, respectively k_r and k_{nr}. A population of N identical molecules excited at $t = 0$ will then follow first-order kinetics. Accordingly, we write

$$\frac{dN(t)}{dt} = -(k_r + k_{nr})N(t) \tag{1}$$

and then $N(t) = N_0 \exp(-(k_r + k_{nr})t)$.

On the other hand, the *fluorescence lifetime* τ of a given fluorescent species is defined as the average time during which an excited molecule will stay excited. Given a population of molecules, it can be written as

$$\tau \stackrel{def}{=} \frac{\int_0^{+\infty} tN(t)dt}{\int_0^{+\infty} N(t)dt}. \tag{2}$$

It follows that $\tau = (k_r + k_{nr})^{-1}$ where k_r is a constant for each fluorescent compound and k_{nr} varies with the environment. In particular, it can be increased by a very short range process between two compounds called FRET [8]. One common application of FLIM is for the measure of FRET, which can be used as an actual indication of *in vivo* proximity between proteins at the nanometer scale.

It is usually not possible to know the number $N(t)$ of excited molecules at a given time. However it is possible to measure the number of fluorescent molecules at a given time (see below), which is the derivative of $N(t)$ times $k_r/(k_r + k_{nr})$, to account for non radiative returns. If $I(t)$ denotes the fluorescence intensity at time t, we have

$$I(t) = A_0 \exp\left(-\frac{t}{\tau_0}\right). \tag{3}$$

Finally, in the experiments, there can be several populations of molecules with different lifetimes at a given point;

as the ordinary differential equation above is linear, the expected number of excited molecules will be a sum of negative exponential whose rates are assumed to be constant:

$$I(t) \;=\; \sum_{k=0}^{K-1} A_k e^{-t/\tau_k} \tag{4}$$

where K is the number of fluorescent species and A_k and τ_k denote the associated initial fluorescence intensity and lifetime respectively.

Measuring FLIM via TCSPC The aim of FLIM microscopy is to measure τ_k and A_k at each spatial position in the image. The values τ_k typically range between 10 picoseconds and 100 nanoseconds, needing highly specialized instrumentation. The first class of methods explicitly records the response of a sample to an excitation pulse, aims at fitting the fluorescence decay with the mono- or multi-exponential models. A second class of methods, which will not be discussed further here, use a modulated excitation (i.e. sinusoidal) and recover the lifetimes from the phase delay of the response [4].

In TCSPC one literally counts each detected photon. A laser pulse is fired repeatedly, then for each pulse either one or zero photon are received and their arrival time from the laser pulse is recorded. Depending on the sample, that operation is repeated for a variable duration and photon arrival times accumulated until the whole decrease can be recovered. An acquisition can take up to several minutes (for 128×128 or 256×256 images) to accumulate enough photons on each pixel to be able to extract a fluorescence decay and thus to derive fluorescence lifetime. For the live cell experiments, we focus on the case where only a few dozens of photons are collected at each pixel (see Fig. 1).

Depending on the detector used, and given the very small time scale, the arrival time is recorded with a finite precision. This imprecision will lead to a convolution of the data with the Instrument Response Function (IRF) f_{IRF}. Thus, taking into account a constant additive background noise b, the full model of the measured intensity is defined as

$$\lambda(t) \;=\; \int f_{IRF}(t' - \delta) I(t - t') dt' \;+\; b \tag{5}$$

where δ is a zero-time shift to be estimated. This value is supposed to be constant and is estimated separately given the depending estimated parameters (see Section 4 for implementation detail). Note that $\lambda(t) = I(t)$ if the background noise b is zero and the IRF is ideal (close to a Dirac distribution).

Moreover, if the sample being studied is a living cell, the (unknown) movements of the proteins will blur the acquired data. The longer the acquisition takes, the more important this effect will be. Averaging the estimation in a neighbourhood as proposed in the next section is a convenient way to address this issue.

3. POINTWISE AND NEIGHBORHOOD-WISE MAXIMUM LIKELIHOOD ESTIMATION

From the point of view of the population of excited molecules, the probability of emitting a photon is quite small, and the population quite large. Thus, according to the law of rare events, the number of emitted photons at a given time is expected to follow a Poisson distribution with parameter $\lambda(t)$ given by (5), that is the average number of photons emitted at a given time. Thus, at a given pixel, we assume that the measured data are realizations of a Poisson process.

Poisson statistics and counts Let $\{y(i,t) \in \mathbb{N}\ i \in \mathcal{S},\ t \in [0 \dots T]\}$ be the noisy observation data (time-dependent photon counts) where $y(i,t) \in \mathbb{N}|$ denotes the observation at pixel i and time t and $\mathcal{S} \subset \mathbb{Z}^2$ is the set of $n = |\mathcal{S}|$ pixels in the image, that is the regular grid of pixels. Moreover, we assume that the $A_k(i)$ and $\tau_k(i)$ values, $k = 0, \dots K-1$, are spatially varying and we denote $\lambda(i,t)$ (see (3)) the expected number of measured photons at pixel i at time t. It follows that

$$y(i,t) \;\sim\; \mathcal{P}_o(y(i,t);\, \lambda(i,t)) \;=\; \frac{(\lambda(i,t))^{y(i,t)} e^{-\lambda(i,t)}}{y(i,t)!} \tag{6}$$

where $\mathcal{P}_o(\cdot;\, \lambda)$ is the Poisson distribution with parameter λ.

Let $\theta(i) = \{A_0(i), \tau_0(i), \dots, A_{K-1}(i), \tau_{K-1}(i)\}, \forall i \in \mathcal{S}$ be the vector of unknown parameters to be estimated at each spatial position. The values $\{\tau_k(i)\}$ represent the parameters of interest. Generally the $\{A_k(i)\}$ values are not interpreted for molecular interaction analysis. Several estimation methods has been proposed but the Maximum Likelihood (ML) estimator is recommended in most cases, leading to a formal criterion to fit the data with the model.

Variance stabilisation and Generalized Anscombe transform is a convenient way of dealing with Poisson variables [2, 5]. However, the number of photons should be high enough if applied. Anscombe transform is also known to be biased, and the bias depend on the number of photons. Accordingly, the bias varies with time as the number of photons exponentially decreases, leading to wrong lifetime estimation. Bias can be theoretically corrected but the precision of the gain parameter is quite small [2, 5]. Finally, we decided to exploit explicit Poisson models in our studies.

Pointwise ML estimation In this section, we focus on the analysis of a individual pixel. In this simple case, the ML estimation of parameter $\theta(i)$ at a given pixel $i \in \mathcal{S}$ is (under the conditional independence hypothesis) defined as:

$$\widehat{\theta}(i) \;=\; \arg\max_{\theta(i)} \prod_{t=0}^{T} f(y(i,t)|\, \theta(i)) \tag{7}$$

where $f(z)$ denotes the probability density function of z. This

Approximate photon counts	20				120			
Patch size (pixels)	0	3	5	7	0	3	5	7
Mean square estimate	3902 ± 5144	2084 ± 234	2282 ± 195	2326 ± 128	2178 ± 372	2266 ± 141	2327 ± 68	2336 ± 53
Poisson ML estimate	2427 ± 747	2337 ± 197	2331 ± 117	2330 ± 83	2356 ± 272	2341 ± 88	2341 ± 52	2340 ± 37

Table 1. Comparison of mean square and Poisson maximum likelihood estimation for a low (120) or very low (20) photon counts and 4-pixel neighbourhood size.

criteria is transformed into a minimization criterion as:

$$\widehat{\theta}(i) \;=\; \arg\min_{\theta(i)} \; -\sum_{t=0}^{T} \log f(y(i,t)|\,\theta(i)). \qquad (8)$$

As explained above, the probability density function should be a Poisson distribution $f(y(i,t)|\,\theta(i)) = \mathcal{P}_o(y(i,t);\,\lambda(i,t))$. It is worth noting that if $\lambda(i,t)$ is high enough $\mathcal{P}_o(\cdot;\,\lambda(i,t))$ tends to the Gaussian distribution with mean and standard deviation $\lambda(i,t)$, i.e. $\mathcal{N}(\lambda(i,t),\lambda(i,t))$. This approximation is actually considered in most methods, including commercial software, and amounts to minimizing a least-square criterion:

$$\widehat{\theta}(i) \;=\; \arg\min_{\theta(i)} \; \sum_{t=0}^{T} \left(\frac{y(i,t)-\lambda(i,t)}{\sigma} \right)^2 \qquad (9)$$

where σ denotes either $y(i,t)$ or $\lambda(i,t)$ [8].

In our approach, we consider the ML estimation in the Poisson case defined as

$$\widehat{\theta}(i) \;=\; \arg\min_{\theta(i)} \; \sum_{t=0}^{T} y(i,t)\log(\lambda(i,t)) - \lambda(i,t) - \log(y(i,t)!). \qquad (10)$$

Since the number of photons does not exceed three photons at a given time t, the minimization turns out to be hard to solve. A convenient way to overcome this difficulty consists in assuming that the lifetime is constant over a small spatial neighborhood, which allows us to summing the photons and then to help to find better minimum. Unfortunately, this method tends to introduce blur in the reconstructed image and the lifetime is not estimated accurately at a given pixel.

Neighborhood-wise ML estimation In this section, we assume that $\theta(i)$ is constant over a region $\mathcal{R} \subset \mathcal{S}$ and the $\{y(i,t)\}, i \in \mathcal{R}$ have the same Poisson distribution. It follows that, under the conditional independence assumption,

$$\widehat{\theta}_{\mathcal{R}}(i) \;=\; \arg\min_{\theta_{\mathcal{R}}(i)} \; -\sum_{j\in\mathcal{R}} \sum_{t=0}^{T} \log f(y(j,t)|\,\theta_{\mathcal{R}}(i)) \qquad (11)$$

where $\widehat{\theta}_{\mathcal{R}}(i), \forall i \in \mathcal{S}$ are the parameters estimated from pixels j taken in the neighborhood $\mathcal{R}(i)$ of pixel i. Conveniently, the region $\mathcal{R}(i)$ is a circular neighborhood $\mathcal{T}(i)$ (space-time tube) centered at pixel i. Small 3×3 patches are recommended but 5×5, 7×7 and 9×9 patches are required if the Poisson distribution is approximated by a Gaussian distribution.

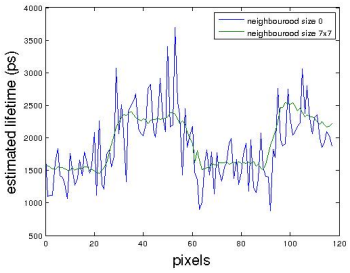

Fig. 3. Test on simulated data with lifetime following a step function with target values of 2340 ps and 1560 ps and about 20 photons per pixel using ML Poisson estimation (only the profile for one line is shown for clarity). A patch size of 1×1 pixel leads to a very noisy estimate while a patch of size of 7 $\times$ 7 pixels leads to blurred results.

Instead of considering a unique neighborhood size for all pixels, we propose to collect a set of temporal signals in a spatially-varying neighborhood. A set of nested space-time tubes is then defined as described in [1]. The optimal diameter of each tube $\mathcal{T}(i)$ at pixel i is found by minimizing the point-wise L_2 risk $\mathbb{E}[(\lambda(i,t) - \widehat{\lambda}(i,t))^2]$ of the estimator $\widehat{\lambda}(i,t) \;=\; \frac{1}{|\mathcal{T}(i)|}\sum_{j\in\mathcal{T}(i)} A_0(j)e^{-t/\tau_0(j)}$ (mono-exponential case), which amounts to balancing the bias an the variance of the estimator [1, 2]. Given an optimal tube $\widehat{\mathcal{T}}(i)$, the parameters $\theta_{\mathcal{R}}(i)$ are estimated according to (11).

4. EXPERIMENTS RESULTS

Our aim is to demonstrate that least-squares and Gaussian models, as used in commercial software, are not appropriate to handle low photon dose in TCSPC FLIM imaging. It turns out that, even when considering a Poisson model, we need to exploit data over a spatial neighborhood for robust estimation.

In our experiments, we used the standard commercial TCSPC FLIM Becker & Hikl software and the proposed least-square algorithm. For the ML Poisson model, optimization is performed using the GNU Scientific Library minimization algorithms. At the initialization stage of the algorithm, the background noise b and the IRF f_{IRF} function with zero-time shift δ are estimated or set by the user. It is worth noting that quality of the solution of the Poisson model is more sensitive to the initialization than the Gaussian (least-squares) model. In practice, a least-square solution is used to initialize the ML Poisson estimation.

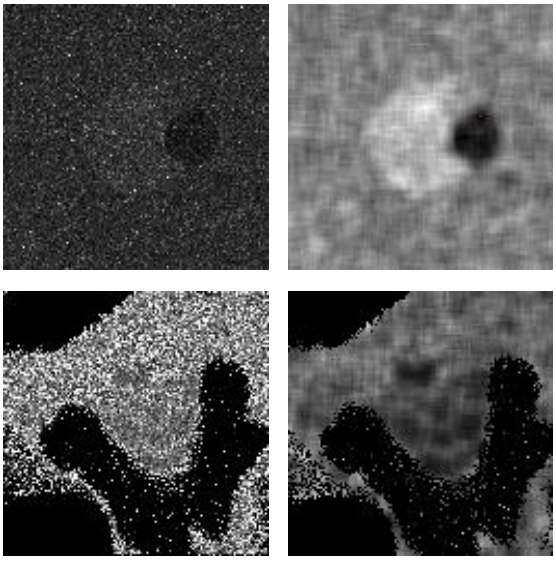

Fig. 4. Test on a more complex simulation (top) and an experimental image (bottom), for patch size of 1×1 (right) and 7×7 (left). A large patch size is needed to deal with the low photon count, which results in a loss of resolution.

Simulations have already been considered to evaluate the performance of lifetime estimation algorithms. Here the model (6) is mainly used for generating artificial data. An image pixel is a sample realization of a Poisson variable with parameter $A_0(i)$ and $\tau_0(i)$ (i.e. $K = 1$). The instrument is assumed to be ideal with a background noise $b = 0$ and we assume an ideal Dirac function to represent IRF. In addition, as observed on real data, 128×128 images are generated with 256 time points and the actual signal begins at $t = 30$. The lifetime τ_0 is taken depending on experiment between 40 and 75 frames, which would correspond to 1560 to 2925 ps with the frame-rate. Different intensities are used corresponding to an average number of photons equal to 120 or 20 per pixel, i.e. photon counts ranging from low to very low photon doses, in agreement with experimental data.

In the first part of experiments, the algorithms were tested on images with a constant and unique lifetime. The fitting accuracy can be assessed by examining the standard deviation of the histogram of the lifetime images. Results are shown in Table 1 and it can be seen that ML estimator under Poisson model outperforms the mean-square estimator for the considered range of photon counts. When the patch size increases, the Gaussian model is more justified and the fitting is improved. Nevertheless, the results are quite poor if we consider 1×1 pixel patches combined with the Poisson model.

Simulating spatially varying lifetime is straightforward, but assessing the fitting accuracy is not an easy task. To illustrate the issue arising with the averaging over large areas, an artificial image depicting a fluorescence lifetime with a step function was generated. Figure 3 shows the profile of the estimated lifetime under Poisson model for two different patch sizes (1×1 patches and 7×7 patches). The expected loss of spatial accuracy is quite demonstrative. A demonstration of the adaptive estimation is currently investigated.

Figure 4 depicts qualitative results on images showing living HeLa cells expressing OCRL1wt protein fused to GFP fluorescent protein and Rab5 in its active (GTP bound) form fused to mRFP1 fluorophore, which shows local variation of lifetime corresponding to molecular proximity revealed by FRET. Considering larger patches is mandatory for Poisson ML estimation, which tends to introduce undesirable blur and spatial inaccuracy.

5. CONCLUSION

In this work we addressed a challenging and original problem in image processing and microscopy. A careful analysis of the biophysical phenomenon and instrument models are used to derive a proper ML framework for lifetime estimation. Evaluation and comparison was shown on simulated and real biological data. It was shown that using spatial information and Poisson statistics is needed to deal with biological application with a very low photon counts. In particular, the application to real biological data allowed us to show the spatial localisation of interactions, a new result which cannot be achieved with commercial methods. Nevertheless, even Poisson ML adapted to the problem result in an appreciable loss of spatial accuracy. The main challenge is now to extend the framework to deal with multi-exponential decay estimate and adaptive neighborhoods [1], a challenge we need to address for a large class of biological studies.

6. REFERENCES

[1] J. Boulanger, C. Kervrann, P. Bouthemy. "Estimation of dynamic background for fluorescence video-microscopy", in Proc. ICIP'06, pp. 2509-2512, Atlanta, USA, 2006.

[2] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, J. Salamero, "Patch-based non-local functional for denoising fluorescence microscopy image sequences." *IEEE Trans. Medical Imaging*, vol. 29, pp. 442–454, 2010.

[3] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, *Plenum Press, New York, London,* 1983.

[4] J.R. Lakowicz and B.R. Masters, "Principles of fluorescence spectroscopy", *J. Biomedical Optics*, vol. 13, pp. 029901, 2008.

[5] M. Makitalo, A. Foi, "Optimal inversion of the generalized anscombe transformation for Poisson-Gaussian noise", *IEEE Trans. Image Processing*, vol. 2, no. 1, pp. 91-103, 2013.

[6] D.V.O. O'Connor and D. Phillips, Time-correlated Single Photon Counting, *Academic Press, London,* 1984.

[7] T. Oida , Y. Sako, A. Kusumi, "Fluorescence lifetime imaging microscopy (flimscopy). Methodology development and application to studies of endosome fusion in single cells". *Biophysical J.*, vol. 64, pp. 676-685, 1993.

[8] S. Pelet and M.J.R. Previte and L.H. Laiho and P.T.C. So, "A fast global fitting algorithm for flurorescence lietime imaigng microscopy based on image segmentation", *Biophysical J.*, vol. 87, no. 4, pp. 2807-2817, 2004.