



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Liu, L; Kan, A; Leckie, C; Hodgkin, PD

Title:

Comparative evaluation of performance measures for shading correction in time-lapse fluorescence microscopy

Date:

2017-04-01

Citation:

Liu, L., Kan, A., Leckie, C. & Hodgkin, P. D. (2017). Comparative evaluation of performance measures for shading correction in time-lapse fluorescence microscopy. *Journal of Microscopy*, 266 (1), pp.15-27. <https://doi.org/10.1111/jmi.12512>.

Persistent Link:

<https://hdl.handle.net/11343/292246>

Comparative evaluation of performance measures for shading correction in time lapse fluorescence microscopy

Liyan Liu*, Andrey Kan[†], Christopher Leckie*, and Philip D. Hodgkin[†]

* *Department of Computing and Information Systems, The University of Melbourne, Australia*

[†] *Division of Immunology, Walter and Eliza Hall Institute of Medical Research, Australia*

Abstract- Time lapse fluorescence microscopy is a valuable technology in cell biology, but it suffers from the inherent problem of intensity inhomogeneity due to uneven illumination or camera nonlinearity, known as shading artefacts. This will lead to inaccurate estimates of single-cell features such as average and total intensity. Numerous shading correction methods have been proposed to remove this effect. In order to compare the performance of different methods, many quantitative performance measures have been developed. However, there is little discussion about which performance measure should be generally applied for evaluation on real data, where the ground truth is absent. In this paper, the state-of-the-art shading correction methods and performance evaluation methods are reviewed. We implement ten popular shading correction methods on two artificial datasets and four real ones. In order to make an objective comparison between those methods, we employ a number of quantitative performance measures. Extensive validation demonstrates that the coefficient of joint variation (CJV) is the most applicable measure in time lapse fluorescence images. Based on this measure, we have proposed a novel shading correction method that performs better compared to well-established methods for a range of real data tested.

Keywords- Time lapse fluorescence microscopy; shading correction methods; quantitative performance measures.

Correspondence to: Liyan Liu, Department of Computing and Information Systems, The University of Melbourne. Address: 156-292 Grattan St, Parkville 3010, Australia. Tel: 61-04-52521296. E-mail: liyanl@student.unimelb.edu.au

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jmi.12512

This article is protected by copyright. All rights reserved.

1 Introduction

Fluorescence microscopy coupled with single cell tracking and an appropriate data analysis method has proved to be an increasingly important tool in cell biology (Dowling et al., 2014). In particular, we can analyze cell behavior over the cell cycle, based on the resulting fluorescence time courses. Thus, it is important to acquire a correct interpretation of the measured fluorescence levels such as cell average intensity and total intensity over time, since they are indicative of the internal cell state (i.e., the amount of the corresponding protein at a given time). Compared to confocal microscopy, time lapse microscopy is better suited for long-term imaging because it exposes the sample to less light. However, it suffers from the inherent problem of intensity inhomogeneity, also known as shading (Likar et al., 2000), or uneven illumination. Generally, there are three types of effects that cause the difference of cell intensities: natural variation between cells, position-dependent acquisition effects (i.e., shading) and time-dependent acquisition effects such as cross-talk (Carlsson & Mossberg, 1992) and photobleaching (Vicente et al., 2007). In order to analyze cell behavior, we assess changes in natural cell fluorescence levels over time. Therefore, these acquisition effects need to be eliminated. In this paper, we are focusing on shading removal.

Because of this shading artefact, pixel intensity of the same tissue varies with its location, which can be seen in Figure 1. This will lead to inaccurate measurements of single-cell features such as average intensity and total intensity. Meanwhile, it has been theoretically proved that shading has an effect on these single-cell features (Coster et al., 2014). As is shown in Figure 2, there is a strong correlation (-0.33) between cell average intensity and cell location for one shaded image. Due to random cell placement, there should be no correlation between cell intensity and cell location. Thus, the value of -0.33 is not negligible since it may cause substantial variation in the average intensities. Moreover, many image processing applications such as segmentation, tracking or quantitative analysis are highly sensitive to shading artefacts. Hence, it can be important to eliminate the shading effect beforehand.

In general, shading could be either object-independent or object-dependent (Tomažević et al., 2002). Object-independent shading is caused by imperfect image acquisition, while object-dependent shading arises from the preparation of ob-

jects by, for example, staining inhomogeneity. A prospective correction method can be applied to address the object-independent shading by using a calibration slide and an acquisition protocol, but these are not always available. Hence, a number of retrospective methods have been developed to correct both object-dependent and object-independent shading when the only available data is the image itself.

In this study, we aim to choose an appropriate correction method from a number of proposed shading correction approaches for real data, where the ground truth is absent. In order to show whether one correction method is feasible for fluorescence images and superior than other methods, numerous quantitative measures have been proposed. However, there is no consensus regarding which performance measure should be applied for evaluation on real data. Furthermore, most shading correction approaches (Belaroussi et al., 2006; Hou, 2006; Vovk et al., 2007; Balafar, 2012) and evaluation strategies (Arnold et al., 2001; Tomažević et al., 2002; Chua et al., 2009) are specially developed for magnetic resonance (MR) images, which may be not suitable for adaptation to time lapse fluorescence microscopy images. Moreover, some correction methods have many parameters to tune, resulting in the difficulty of parameter setting. Consequently, it is not easy to choose an optimal approach to deal with shading in time lapse fluorescence microscopy. Therefore, our purpose is to determine a quantitative measure for performance evaluation on real data.

In summary, our contributions are as follows: 1) We comprehensively review the state-of-the-art shading correction methods and performance evaluation approaches (Section 2). 2) We address the problem of performance evaluation on real data, in which the ground truth is not available (Section 3). 3) We compare and evaluate the existing popular shading correction methods by conducting experiments on synthetic and real datasets (Section 3.5). 4) Furthermore, a novel shading correction method based on CJV has been developed (Section 3.4).

2 Shading Correction Methodology

In this section, we first formulate the problem by addressing the shading effect that needs to be considered in the following analysis. Then we provide an overview of the most popular shading correction methods in the literature. Finally, we summarize different performance evaluation approaches in the field of shading correction.

2.1 Problem Formulation

Time lapse fluorescence microscopy adopts a multi-channel imaging mechanism to record spatiotemporal protein accumulation dynamics within cells. Due to nonuniform illumination, the intensity varies within an image, and from image to image. The relationship between the acquired nonuniform image and the shading-free image could be formulated as (Likar et al., 2000)

$$N = S_M \odot U + S_A \quad (1)$$

where N represents the nonuniform image, U indicates the real uniform image, while S_M and S_A denote the multiplicative shading component and additive shading component, respectively. The notation $\odot$ refers to point-wise multiplication.

The goal of shading correction is to estimate both shading components from the acquired nonuniform image. In reality, this problem could be further simplified by assuming that only one shading component degrades the uniform image. Thus, the relation between the nonuniform image and the uniform image could be expressed by a multiplicative or an additive model

$$N = S_M \odot U \quad (2)$$

$$N = U + S_A \quad (3)$$

where U could be reformulated by the sum of a biologically relevant foreground fluorescence from cells F and a background fluorescence B

$$U = F + B \quad (4)$$

In summary, the image formation model could be classified into three categories, namely multiplicative and additive, multiplicative only and additive only. For simplicity, most existing proposed shading correction methods are based on either multiplicative or additive models. Regardless of the model used in shading correction methods, our goal is to find an optimal one from numerous shading removal approaches by performing sound evaluation on fluorescence images.

2.2 Shading Correction Methods

Retrospective shading correction methods make few assumptions about the acquisition process and mainly rely on the information from the acquired images. Existing popular retrospective approaches include surface fitting (SF) (Russ, 2011), rolling ball algorithm (RB) (Sternberg, 1983), signal envelope estimation (SE) (Reyes-Aldasoro, 2009), homomorphic filtering (HF) (Gonzalez & Woods, 2002), gradient distribution (GD) (Zheng et al., 2009), entropy minimization (EM) (Likar et al., 2000), information minimization (IM) (Likar et al., 2001), non-parametric non-uniform intensity normalization (N3) (Sled et al., 1998) and an improved N3 correction method called N4ITK (Tustison et al., 2010). As summarized in Table 1, those methods adopt different image formation models, rely on different assumptions, include different number of tuning parameters, and have different advantages and disadvantages.

The SF algorithm applies least-squares fitting of a polynomial or spline to estimate the shading component by selecting a number of points on the background manually or automatically. Manual selection is subjective and time-consuming, whereas automatic selection assumes good global support of the background, which is not always true. The idea of RB is to use morphological opening of a gray-scale image with a sphere (also called a top hat filter) to obtain an estimation of the background. Here the assumption is that the size of the objects of interest is smaller than the background variation. SE removes the shading component by estimating the envelope of the image through iterative stretching. An implicit assumption involved here is that the objects are either all darker or all brighter than the background.

The fourth correction method HF assumes that the shading is the low-frequency content that could be separated from the high-frequency spectrum of the true image. In some cases, due to the nature of the image, the spectra of shading and shading-free data cannot be separated, which makes it difficult to model the shading effect using low frequencies only. Thus, the accuracy of the homomorphic filtering not only depends on the selection of an appropriate high-pass filter but also the nature of the image. A drawback of the HF method is that it will introduce halo artefacts on the boundaries of the image if optimal parameters are not obtained. GD eliminates the shading component by assuming that the image gradient probability distribution is sparse, which is not always the case. This method is computationally expensive and relies on the parameters of the weighting function.

EM estimates the shading-free image by minimizing its entropy because it is

assumed that shading leads to higher entropy. However, it suffers from a fundamental limitation as it encourages local optima in the optimization. In comparison with EM, IM also minimizes the entropy but estimates the inverse of the shading components using a combination of basis functions while the global intensity statistic is preserved. N3 was originally proposed to eradicate the multiplicative shading component in MR images by maximizing the frequency content of the intensity distribution. The corresponding smooth shading field is supposed to be multiplicative and estimated by the B-spline approximation. Compared to N3, N4ITK couples a robust B-spline approximation algorithm with an advantageous optimization strategy for the shading field estimation. Both methods assume that the true image intensities are independent identically distributed random variables, whereas it is not the case for fluorescence images of cells as the intensity varies a lot within cells. Moreover, the N3 algorithm may converge to local minima in the objective function and the N4ITK method is extremely computational expensive. In particular, they require substantial manual work for parameter setting when applied to time lapse fluorescence images.

2.3 Performance Evaluation Strategies

As mentioned above, there are a large number of measures for comparing the quality of shading correction. The evaluation methods can be categorized into qualitative and quantitative methods. The former is a subjective analysis, which compares the histogram or the intensity profiles before and after shading correction, while the latter adopts some numerical measure as an objective comparison. Hence, we focus on the quantitative evaluation, which could be further divided into two major categories, namely, reference based and reference-free based evaluation. Here reference means the true shading-free image or true shading component. In the following analysis, we list and compare some popular quantitative performance measures for evaluation.

2.3.1 Reference based evaluation

When a reference is available, four quantitative evaluation measures are widely used: normalized L_2 -norm (Chua et al., 2009), correlation coefficient, root mean squared error (RMSE) and peak signal-to-noise ratio (PSNR) (Babaloukas et al., 2011). The normalized L_2 -norm calculates the difference between the true shading-

free image/shading field and the estimated one expressed by

$$L_2 = \min_w \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N (w I_{est}(i, j) - I_{true}(i, j))^2}{\sum_{i=1}^M \sum_{j=1}^N (I_{true}(i, j))^2}} \quad (5)$$

where w is the normalization coefficient since shading correction can lead to an arbitrary scaling of the reference image. I_{est} and I_{true} with size $M \times N$ represent the estimated shading-free image/shading field and the true shading-free image/shading field, respectively.

RMSE is used to measure the difference between the true and the estimated shading-free image/shading field defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N (I_{est}(i, j) - I_{true}(i, j))^2}{M \times N}} \quad (6)$$

Spearman's correlation coefficient (r) is adopted to show the similarity between the estimated image and the true one. PSNR is the ratio of the maximum possible pixel value $Max_{I_{est}}$ in an image to the root mean squared error $RMSE$

$$PSNR = 20 \log \left(\frac{Max_{I_{est}}}{RMSE} \right) \quad (7)$$

2.3.2 Reference-free based evaluation

This evaluation method is applied in the absence of reference images. The shading correction results can be quantitatively assessed by estimating the intensity variation before and after correction. More specifically, most commonly used quantitative measures include standard deviation (σ) for the whole image, coefficient of variation (CV) (Vovk et al., 2007), coefficient of joint variation (CJV) (Vovk et al., 2007) and signal-to-noise ratio (SNR).

CV is defined as a ratio of standard deviation σ to mean value μ of one tissue class (C), namely foreground or background in an image

$$CV = \frac{\sigma(C)}{\mu(C)} \quad (8)$$

One drawback of CV is that the mean value but not the variance of the class is changed when the uniform additive intensity transformation occurs. Another drawback is that it is calculated for a single class, which makes it difficult to

assess the overall correction quality for an image composed of many classes, particularly when CV improves for one class but not for the others. In order to avoid those drawbacks, for an image composed of two classes (C_1 and C_2 , e.g., cell and background), CJV has been proposed by estimating the overlap between them

$$CJV = \frac{\sigma(C_1) + \sigma(C_2)}{|\mu(C_1) - \mu(C_2)|} \quad (9)$$

CJV is invariant to the uniform additive and multiplicative transformation, but the disadvantage is that we need to perform segmentation to obtain different classes in an image. Furthermore, it has been shown that the evaluation results depend on the accuracy of the segmentation (Chua et al., 2009). Because true shading-free images are unlikely to be available for real data, CV and CJV could not provide information on the amount of shading remaining after correction. In reality, we aim to obtain a sound segmentation result.

As summarized in Table 2, the disadvantage of reference based evaluation is that the true shading-free images or shading components have to be known, which is not feasible for real data. In particular, CV and CJV require segmentations as a preprocessing step before evaluation. Even though reference-free based measures could be applied for time lapse fluorescence images, there is no consensus regarding which performance measure should be adopted. Thus, our purpose is to determine a measure that could be generally used.

3 Experimental Results

In the absence of ground truth for real datasets, our quality evaluation strategy is to determine one reasonable quantitative measure based on artificial datasets. In this case, we first generate artificial datasets that are representative of real datasets used in the experiments. Then we determine a reference-free performance measure by comparing the consistency between one reference-based baseline measure and all the others. Subsequently, we proposed a novel shading correction method, which aims to minimize the CJV value of the estimated shading-free image. Next we explicitly describe the implementation details of the existing popular shading correction methods. By comparing and evaluating all the above-mentioned shading correction methods, we conclude an optimal approach to remove shading for real fluorescence images by the determined measure.

3.1 Tested Datasets

In this study, the proposed shading correction strategies are tested on both artificial datasets and true fluorescence images. The real datasets were sampled from previously acquired multi-channel videos (Liu et al., 2014; Kan et al., 2011). In brief, CpG-stimulated B lymphocytes were placed in physically separated compartments called wells at low density (Day et al., 2009) and continuously imaged for several days (Dowling et al., 2014). The role of the well is to prevent the cells from migrating outside the field of view. Ubiquitin-GFP mice were used in some videos, providing the fluorescence green signal (Kan et al., 2011). In other videos, fluorescence ubiquitination cell cycle indicator (FUCCI) mice were used (Liu et al., 2014), in which case red and green fluorescence signals were recorded. In all cases, an image sequence in each channel was captured. It is convenient to refer to image sequences using well shapes (i.e., rectangular, hexagonal and square in Figure 3), which, however, do not have any biological significance in these experiments.

We have formed four real datasets reflecting the variability in experimental conditions. For each dataset, we randomly selected ten fluorescent images from one of the video sequences, as is presented in Figure 4. Specifically, dataset 1 and dataset 2 contain images sampled from the video with rectangular wells in the FUCCI red channel and FUCCI green channel, respectively. However, dataset 3 and dataset 4 consist of images sampled from the video with hexagonal wells and square wells in the Ubiquitin-GFP channel, respectively. Due to the low contrast between the background and cells (white dots) in Figure 4, the wells are barely visible for images in the fluorescent channels compared to images in the transmission channels. The difficulty of quality evaluation on real datasets lies in the lack of ground truth: shading-free images. An indirect way is to determine a reasonable quantitative measure from artificial datasets, which is then applied in real datasets.

The synthetic images are generated using a previously published simulation algorithm with some modifications (Lehmussola et al., 2007). This image simulation algorithm is also used in (Peng et al., 2014). Gaussian noise δ is added into the simulated images consisting of two classes: background and cells. In practice, we make the background follow a uniform nonnegative distribution. The intensities in the background and cells vary. The illumination field S is simulated by the second-order polynomial with fixed parameters, and then applied to corrupt the simulated shading-free image by a multiplicative model rather than an additive one. The relationship between the nonuniform image N and the shading-free

image U is given by

$$N = S \odot U + \delta \quad (10)$$

We generated ten synthetic shading-free images, ten illumination images and ten resulting nonuniform images (1040×1388 pixels, non-normalized intensity between 0 and 65535) simultaneously. In this case, we can adopt reference-based quantitative measures to perform evaluation for synthetic datasets. Two artificial datasets are created by setting different parameters of the above-mentioned simulation algorithm. In order to ensure these datasets are representative of real data, we first aim to obtain appropriate parameters in this algorithm by making the normalized empirical cumulative probability distribution of the artificial datasets follow that of the real datasets, as is presented in Figure 5. Then we exploit those parameters to produce the artificial datasets. Figure 6 shows the shading-free image, corresponding shading field and resulting shaded image from artificial dataset 1 and artificial dataset 2. Next we introduce the implementation details of all the above-mentioned shading correction methods in the experiment.

3.2 Implementation Details

For the shading correction methods that require parameter tuning, it is essential to find the appropriate parameter values of this method in order to obtain accurate correction results. As is described in Table 3, the implementation details of the above nine shading correction methods are summarized and compared regarding the tuning parameters, the application tools and the adopted image formation model. All of those algorithms are open source in Fiji (Schindelin et al., 2012), Matlab or MIPAV (McAuliffe et al., 2001).

In the SF method, the shading field is obtained by fitting a parametric polynomial. Some methods such as “shading correction”, “polynomial_fit”, “fit_polynomial” and “nonuniform background removal”¹ embedded in Fiji are widely used (Kuru, 2014). Here we choose the most accurate and least time-consuming algorithm “polynomial_fit”² to fit an Nth order Legendre polynomial, which is then subtracted from the shaded image. Since the shading is assumed to be additive, this algorithm is renamed SFA. In this case, the tuning parameters are the order of the polynomial in the x direction $xOrder$ and y direction $yOrder$. The RB algorithm is performed under Fiji’s function “Subtract Background”, which estimates the

¹http://imagejdocu.tudor.lu/doku.php?id=howto:working:how_to_correct_background_illumination_in_brightfield_microscopy

²http://www.optinav.info/Polynomial_Fit.htm

background by a sphere structuring element. The tuning parameter is the size of this structuring sphere, also called the rolling ball radius R . This value should be at least as large as the radius of the largest object that is not part of the background.

The idea of the SE algorithm is to obtain the shading component through iterative stretching of a flexible surface until it satisfies the optimization criterion. This algorithm is implemented in Matlab, which requires the number of iterations M and critical value δ for the optimization criterion to be adjusted. In the HF algorithm, filtering is performed by convolving the image with a high-pass filter in the frequency domain. It requires fine tuning of four parameters: low frequency bound γ_L , high frequency bound γ_H , constant focus distance D_0 and constant c used to control the transition of the filter. We implement this algorithm called “Homomorphic Filter”³ in Fiji. In this paper, the SE and HF algorithms are extended to work with both 8-bit and 16-bit grayscale images.

Assuming that the image gradient follows a sparse distribution expressed by an exponential function with parameter α , GD deals with shading by optimizing the objective function including such parameters as the order of polynomial D , number of iterations M and smoothness factor ϵ . In order to increase the speed of the GD algorithm, the image is first scaled to a smaller size using subsampling factor s . This algorithm is extended and thus works for both 8-bit and 16-bit grayscale images. In the EM method, the correction is performed by minimizing the entropy after initializing five parameters $a_1 - a_5$ in the additive shading field and five parameters $m_1 - m_5$ in the multiplicative shading field. The former algorithm is implemented in Matlab and the latter in Fiji. The algorithm “Entropy Minimization” in the MIPAV software is a combination of IM and EM, which is renamed IEM here. No parameters need to be tuned for this algorithm except the selection of image models: multiplicative or additive and multiplicative.

Another algorithm in MIPAV is inhomogeneity N3 correction, an iterative approach that estimates the multiplicative shading field by setting seven parameters: 1) threshold θ used to differentiate background and foreground; 2) maximum number of iterations M ; 3) converge threshold β between two iterations; 4) field distance F over which the shading field varies; 5) subsampling factor s ; 6) full width at half maximum of deconvolution kernel φ ; 7) Wiener filter noise n . Next, N4ITK is publicly available through the Insight Toolkit of the National Institutes of Health⁴ and also built in MIPAV. Six parameters are required to be tuned: 1) maximum number of iterations M ; 2) convergence threshold β between two iter-

³<http://svg.dmi.unict.it/iplab/imagej/Plugins/Forensics/homomorphic/homomorphic.htm>

⁴<http://www.ncbi.nlm.nih.gov/pubmed/20378467>

ations; 3) full width at half maximum of deconvolution kernel φ ; 4) Wiener filter noise n ; 5) B-spline mesh resolution level r ; 6) number of control points N in B-spline.

In order to implement all the methods optimally for each dataset, we define the parameters by conducting extensive parameter tuning using grid search before evaluation. In Table 4, we determine the values of the tuning parameters for two artificial datasets and four real datasets. For the N3 method, the values for θ and F are set automatically: minimum value Min of current shaded image plus one and default. Both multiplicative and additive shading fields are considered for IEM algorithms, so we rename it IEMMA. In Table 4, symbol $-$ means that the parameters are fixed in the algorithms and do not need to be adjusted.

3.3 Performance Measure Decision

Due to the lack of ground truth for real datasets, we aim to determine the most appropriate reference-free based measure from artificial datasets and then apply this to the real datasets. Because of the segmentation-based performance measures, we need to conduct segmentation for each image from the real datasets, the results of which are further utilized for the segmentation-based shading correction method. Not every segmentation algorithm requires prior shading correction. We applied "TrackMate" (Tinevez, et al., 2016) in Fiji for real datasets 1 and 2, and automated thresholding in Fiji for real datasets 3 and 4. In all cases, we manually validated that the segmentation results were of appropriate quality. For the artificial datasets, we could achieve the mask images directly during simulation since the ground truth is available. Our main aim is shading correction for the purposes of accurate measurement of cell fluorescence signals over time. Next, we conduct shading correction on two synthetic datasets including ten images each, and evaluate the quality using most proposed measures.

For each method, we calculate the mean value of each measure and then rank all the methods from the best to the worst. Table 6 shows the rank of all the shading correction methods based on the values of the corresponding measures for artificial dataset 1 only. Rank "1" indicates the best quality among all, while "9" the worst. This table shows that it is difficult to determine an optimal approach because the rank varies with performance measures. Here $_c$ and $_b$ refer to the measure from cell and background, respectively. Furthermore, $\overline{CV(\%)}$ and $\overline{\sigma(\%)}$ indicate the average of the relative change of coefficient variation (CV) and standard deviation (σ) from those two classes, respectively. The rank varies with measures, resulting in no consensus on which one should be used. Compared

to reference-free measures such as RMSE and PSNR, the normalized L_2 -norm is resistant to scaling. Thus, we compare the consistency between other measures and L_2 -norm on artificial datasets using Kendall's tau coefficient (Kendall & Gibbons, 1990) and Spearman's correlation coefficient, both of which reach relatively higher values for two artificial datasets on average (0.639, 0.808) when the CJV measure is used. This implies that the CJV based measure is strongly consistent with L_2 -norm.

3.4 Proposed Shading Correction Method

In order to remove shading effects, the existing popular approaches take advantage of different optimization criteria, e.g., minimizing the entropy or gradient of the corrected image. While there are many performance measures that focus on variation in pixel intensity, there are no methods that minimize the variance explicitly. However, we have proved that the CJV-based measure could be widely used for performance evaluation on real data. Therefore, we develop a novel shading correction method that aims to minimize the CJV value of the estimated uniform image. In order to maintain the global intensity characteristics of the images, we model the smoothly varying shading fields as the same expression in (Likar et al., 2000)

$$S_M(x, y) = 1 + m_1x + m_2y + m_3xy + m_4(x^2 - W^2/12) + m_5(y^2 - H^2/12) \quad (11)$$

$$S_A(x, y) = a_1x + a_2y + a_3xy + a_4(x^2 - W^2/12) + a_5(y^2 - H^2/12) \quad (12)$$

where x and y represent the pixel position in an image, and W and H refer to the image width and height, respectively. Considering both multiplicative and additive shading components, the estimated uniform image could be expressed by

$$\hat{U}(x, y) = \frac{N(x, y) - S_A(x, y)}{S_M(x, y)} \quad (13)$$

The idea of the novel shading correction method is to minimize the CJV value of the estimated uniform image by searching for the optimal additive parameters $\mathbf{a}$ and multiplicative parameters $\mathbf{m}$

$$CJV_{min} = \min_{\{\mathbf{a}, \mathbf{m}\}} CJV(\hat{U}(x, y)) \quad (14)$$

This novel correction method is called CJVMA when both shading components are considered, otherwise called CJVM or CJVA. We must perform segmentation for the acquired images beforehand. For fluorescence images that are seriously corrupted, it is difficult and time-consuming to obtain a good segmentation result. Thus, we aim to conduct rough segmentation and achieve relatively good quality. The advantage of CJVMA method is that no parameters need to be tuned. It is a built-in plugin called "CJV Minimization" in Fiji. The source code and the tested datasets mentioned above are publicly available at the website ⁵.

3.5 Performance Evaluation Results

Quantitative evaluation is essential for the objective comparison of the results obtained by different shading correction methods. We conducted experiments on four real datasets by applying the above ten shading correction methods and perform evaluation based on the CJV value. We implement the CJVMA algorithm directly without manual tuning of parameters. Image noise is removed by "Gaussian Blur" in Fiji beforehand. In order to show whether the shading is efficiently removed, we apply the relative change of CJV to the comparative performance evaluation, which is defined as

$$CJV(\%) = \frac{CJV_o - CJV_c}{CJV_o} \times 100\% \quad (15)$$

Here CJV_o and CJV_c refer to the CJV value of the original image and the corrected image, respectively. If one shading correction method succeeds in reducing shading, the relative change $CJV(\%)$ becomes positive. Hence, larger positive values imply better performance. The relative changes of CJV for the above-mentioned methods are given in the form of box-whiskers diagrams, illustrated in Figure 6. Each diagram shows the distribution of relative changes for 10 randomly selected real images. Since the CJVMA method aims to minimize the value of CJV directly, the relative change of CJV is the biggest for these four real datasets. However, RB fails to correct the real dataset 1 because the size of the object of interest is bigger than the background variation. EM and IEMMA do not work well for real dataset 1 in that a higher-order polynomial is required to estimate the shading field with a small local variation. Above all, we have the following conclusions:

⁵<https://github.com/LilyCrystal/CJVMinimization>

- The CJV measure can be generally applied in the performance evaluation on real data, where the ground truth is absent.
- CJVMA performs best among all the existing shading correction methods for those four real datasets.
- In general, CJVMA has a comparative advantage over all the abovementioned shading correction methods. The superiority lies in direct correction without parameter tuning, which saves time and manual work. However, the drawback is that segmentation is required beforehand.

4 Conclusion and Discussion

One of our aims is to analyze cell behavior over the cell cycle by extracting fluorescence time courses from time lapse microscopy. However, the imaging results can be influenced by the inherent problem of the acquisition effects such as cross-talk, autofluorescence (Monici, 2005), photobleaching and shading. In order to achieve a correct interpretation of observed fluorescence levels, it is essential to address these effects before data analysis. We have previously addressed the problem of cross-talk and autofluorescence (Liu et al., 2014). In this paper, we aim to remove the shading effect to avoid inaccurate intensity measurement for the subsequent quantitative analysis. Moreover, we deal with the problem of evaluation on real data without ground truth. In the aspect of performance evaluation, there is no consensus on which quantitative measure could be generally utilized in time lapse fluorescence images. Particularly, some developed shading correction methods and performance evaluation strategies are proposed for the MR imaging community, where the image formation model is different from microscopic images. One of our contributions is that we identified that CJV is more reliable compared to the other measures. Hence, we have developed a novel shading correction method called CJVMA, which shows superior performance compared to the well-established methods for most datasets. Even though a prior segmentation is required, there is no need to adjust the parameters in this algorithm. Another contribution is that we compare and evaluate state-of-the-art correction approaches on real datasets, and identify the most appropriate method. Now all the acquisition effects are analysed, so our future work will be to obtain the real fluorescence levels for real datasets by removing all of them. Since CJVMA algorithm is computational expensive, we also need to design a more efficient correction method.

Acknowledgment

This research has been funded in part by the Chinese Scholarship Council.

References

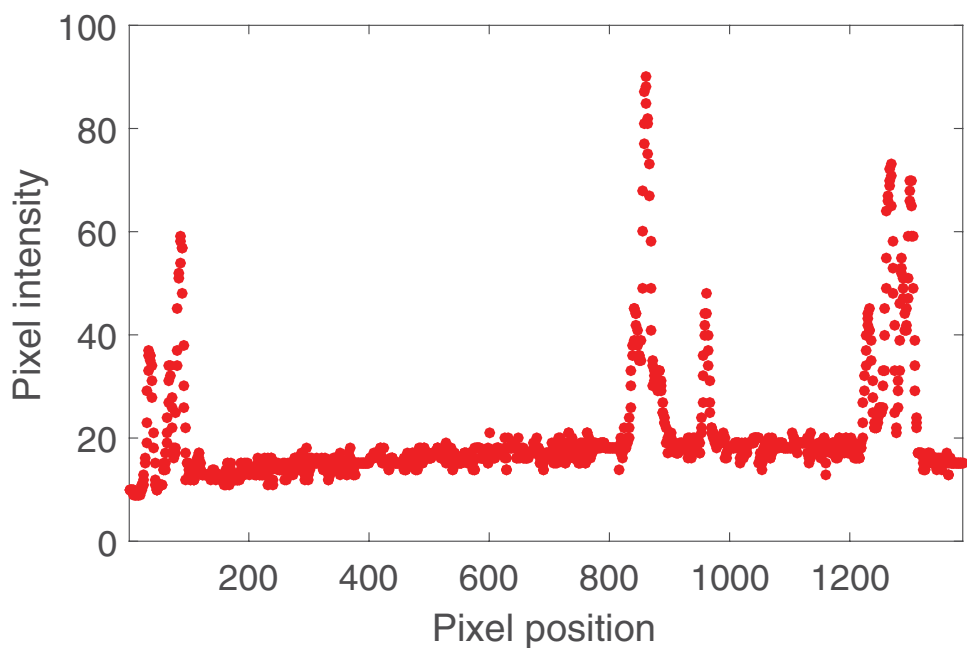
- Arnold, J. B., Liow, J. S., Schaper, K. A., et al. (2001) Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *Neuroimage* **13**(5), 931–943.
- Babaloukas, G., Tentolouris, N., Liatis, S., Sklavounou, A. & Perrea, D. (2011) Evaluation of three methods for retrospective correction of vignetting on medical microscopy images utilizing two open source software tools. *J. Microsc.* **244**(3), 320–324.
- Balafar, M. A. (2012) Review of intensity inhomogeneity correction methods for brain MRI Images. *IJTPE* **4**(4), 60–66.
- Belaroussi, B., Milles, J., Carme, S., Zhu, Y. M., & Benoit-Cattin, H. (2006) Intensity non-uniformity correction in MRI: existing methods and their validation. *Med. Image. Anal.* **10**(2), 234–246.
- Carlsson, K. & Mossberg, K. (1992) Reduction of cross-talk between fluorescent labels in scanning laser microscopy. *J. Microsc.* **167**(1), 23–37.
- Chua, Z. Y., Zheng, W., Chee, M. W. & Zagorodnov, V. (2009) Evaluation of performance metrics for bias field correction in MR brain images. *J. Magn. Reson. Imaging* **29**(6), 1271–1279.
- Coster, A. D., Wichaidit, C., Rajaram, S., Altschuler, S. J. & Wu, L. F. (2014) A simple image correction method for high-throughput microscopy. *Nat. Methods* **11**(6), 602–602.
- Day, D., Pham, K., Ludford-Menting, M. J., Oliaro, J., Izon, D., Russell, S. M. & Gu, M. (2009) A method for prolonged imaging of motile lymphocytes. *Immunol. Cell Biol.* **87**(2), 154–158.
- Dowling, M. R., Kan, A., Heinzl, S., et al. (2014) Stretched cell cycle model for proliferating lymphocytes. *Proc. Natl. Acad. Sci. U.S.A.* **111**(17), 6377–6382.
- Gonzalez, R. C. & Woods, R. E. (2002) Image enhancement in the frequency domain. *Digital Image Processing*, pp. 191–194. 2nd edn. Prentice Hall, New Jersey.
- Hou, Z. (2006) A review on MR image intensity inhomogeneity correction. *Int. J. Biomed. Imaging* **2006**, 1–11.
- Kan, A., Chakravorty, R., Bailey, J., Leckie, C., Markham, J. & Dowling, M. (2011) Automated and semi-automated cell tracking: addressing portability challenges. *J. Microsc.* **244**(2), 194–213.
- Kendall M. G. & Gibbons J. D. (1990) *Rank Correlation Methods*, 5th edn. Griffin, London.

- Kuru, K. (2014) Optimization and enhancement of H&E stained microscopical images by applying bilinear interpolation method on lab color mode. *Theor. Biol. Med. Modell.* **11**(1), 9.
- Lehmussola, A., Ruusuvaori, P., Selinummi, J., Huttunen, H. & Yli-Harja, O. (2007) Computational framework for simulating fluorescence microscope images with cell populations. *IEEE Trans. Med. Imaging* **26**(7), 1010–1016.
- Likar, B., Maintz, J. A., Viergever, M. A. & Pernuš, F. (2000) Retrospective shading correction based on entropy minimization. *J. Microsc.* **197**(3), 285–295.
- Likar, B., Viergever, M. & Pernuš, F. (2001) Retrospective correction of MR intensity inhomogeneity by information minimization. *IEEE Trans. Med. Imaging* **20**(12), 1398–1410.
- Liu, L., Kan, A., Zhou, J. H., Markham, J. F., Hodgkin, P. D. & Leckie, C. (2014) Evaluating spectral unmixing quality in the absence of reference objects for time lapse fluorescence microscopy. *IEEE International Conference on Data Mining Workshop*, pp. 968–975.
- McAuliffe, M. J., Lalonde, F. M., McGarry, D., Gandler, W., Csaky, K. & Trus, B. L. (2001) Medical image processing, analysis and visualization in clinical research. In *Proceedings of the 14th IEEE Symposium on Computer-based Medical Systems*, pp. 381–386.
- Monici, M. (2005) Cell and tissue autofluorescence research and diagnostic applications. *Biotechnol. Annu. Rev.* **11**, 227–256.
- Peng, T., Wang, L., Bayer, C., Conjeti, S., Baust, M. & Navab, N. (2014) Shading correction for whole slide image using low rank and sparse decomposition. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 33–40.
- Reyes–Aldasoro, C. C. (2009) Retrospective shading correction algorithm based on signal envelope estimation. *Electron. Lett.* **45**(9), 454–456.
- Russ, J.C. (2011) Correcting imaging defects. *The Image Processing Handbook*, pp. 235–238. 6th edn. CRC Press, Florida.
- Schindelin, J., Arganda-Carreras, I., Frise, E., et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**(7), 676–682.
- Sled, J. G., Zijdenbos, A. P. & Evans, A. C. (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* **17**(1), 87–97.
- Sternberg, S. R. (1983) Biomedical image processing. *IEEE Comput.* **1**(16), 22–34.
- Tinevez, J. Y., Perry, N., Schindelin, J., et al. (2016) TrackMate: an open and extensible platform for single-particle tracking. *Methods*, 1046–2023.

- Tomažević, D., Likar, B. & Pernuš, F. (2002) Comparative evaluation of retrospective shading correction methods. *J. Microsc.* **208**(3), 212–223.
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A. & Gee, J. C. (2010) N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**(6), 1310–1320.
- Vicente, N. B., Zamboni, J. E. D., Adur, J. F., Paravani, E. V. & Casco, V. H. (2007) Photobleaching correction in fluorescence microscopy images. *J. Phys.: Conf. Ser.* **90**(1), 012068.
- Vovk, U., Pernuš, F. & Likar, B. (2007) A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans. Med. Imaging* **26**(3), 405–421.
- Zheng, Y., Grossman, M., Awate, S. P. & Gee, J. C. (2009) Automatic correction of intensity nonuniformity from sparseness of gradient distribution in medical images. *In Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part II*, pp. 852–859.



(A)



(B)

Figure 1: Shading effect on pixel intensity: (a) one fluorescence image captured from time lapse fluorescence microscopy; (b) intensity profile showing the intensity variation along the red straight line in (a). The data points with relatively bigger pixel intensities (spikes) represent the signal from cells, otherwise from background. This profile indicates that the cell pixel values near the center are elevated more than values near the corners due to uneven illumination. This is not expected in reality as cells are seeded at random.

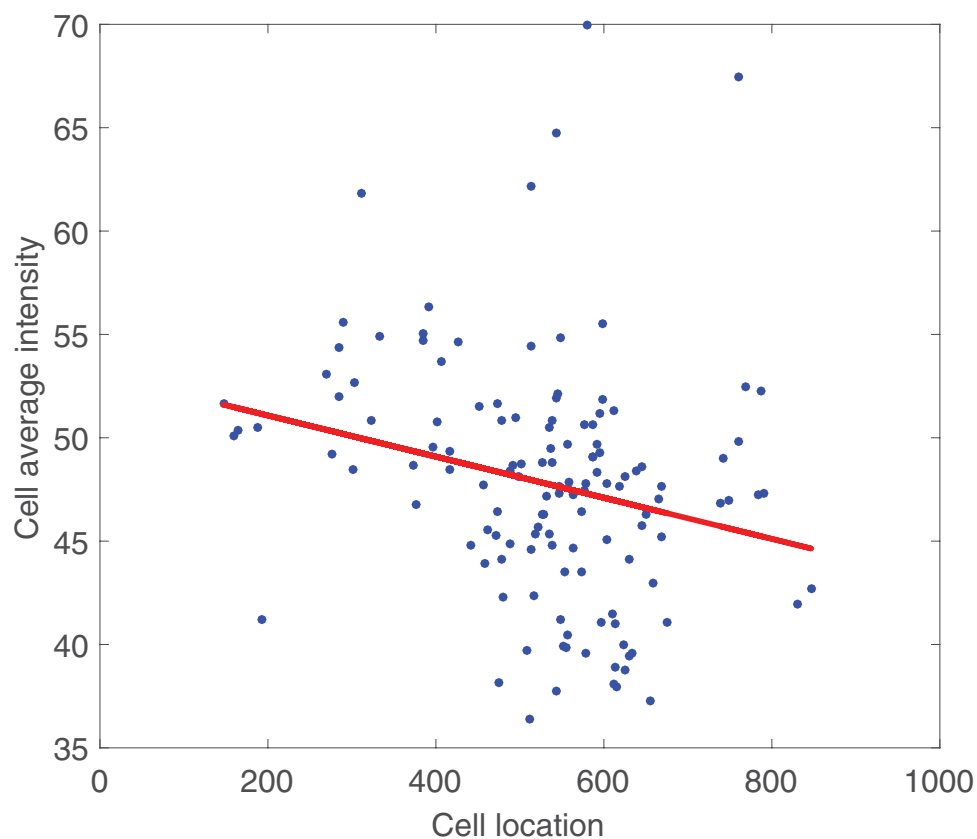


Figure 2: Shading effect on cell average intensity of one shaded fluorescence image. Cell location refers to the distance from the cell centroid to the image centre. The red regression line shows the cell average intensity dependence on the cell location in the image.

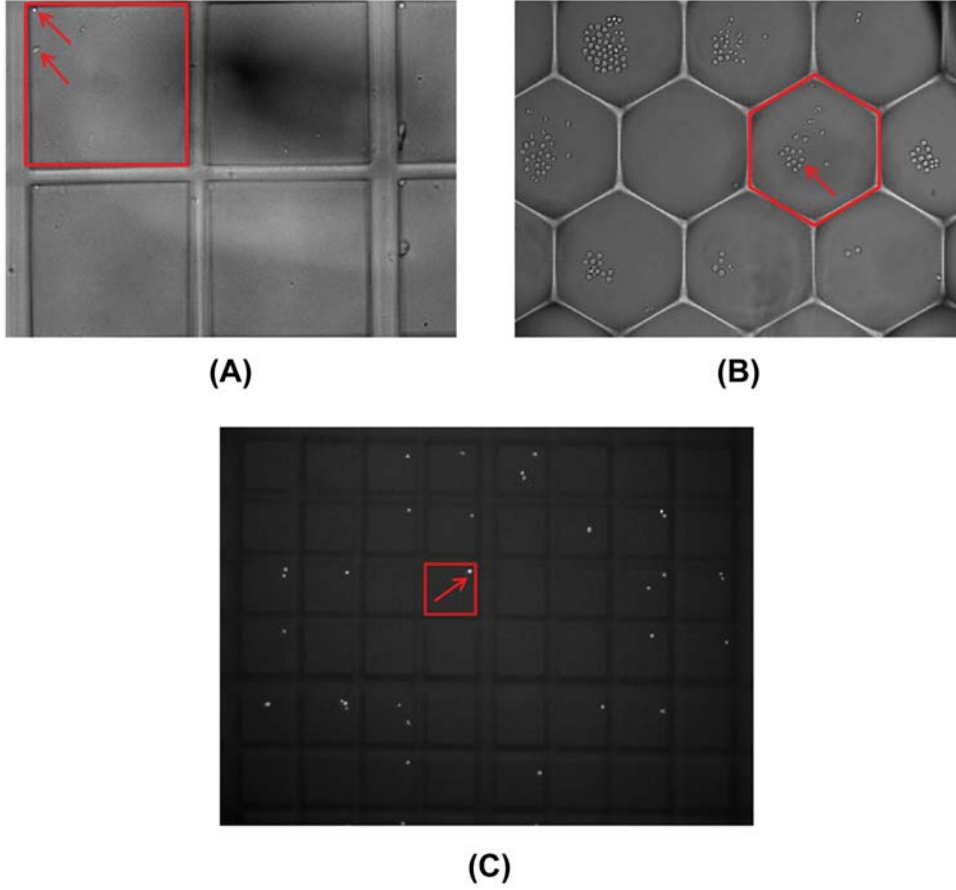


Figure 3: One image from the video where the cells are allocated in: (a) rectangular well, (b) hexagonal well, (c) square well. The first two images are extracted from the transmission out of focus channel, while the last one from the Ubiquitin-GFP channel. Several examples of cells are pointed to by arrows.

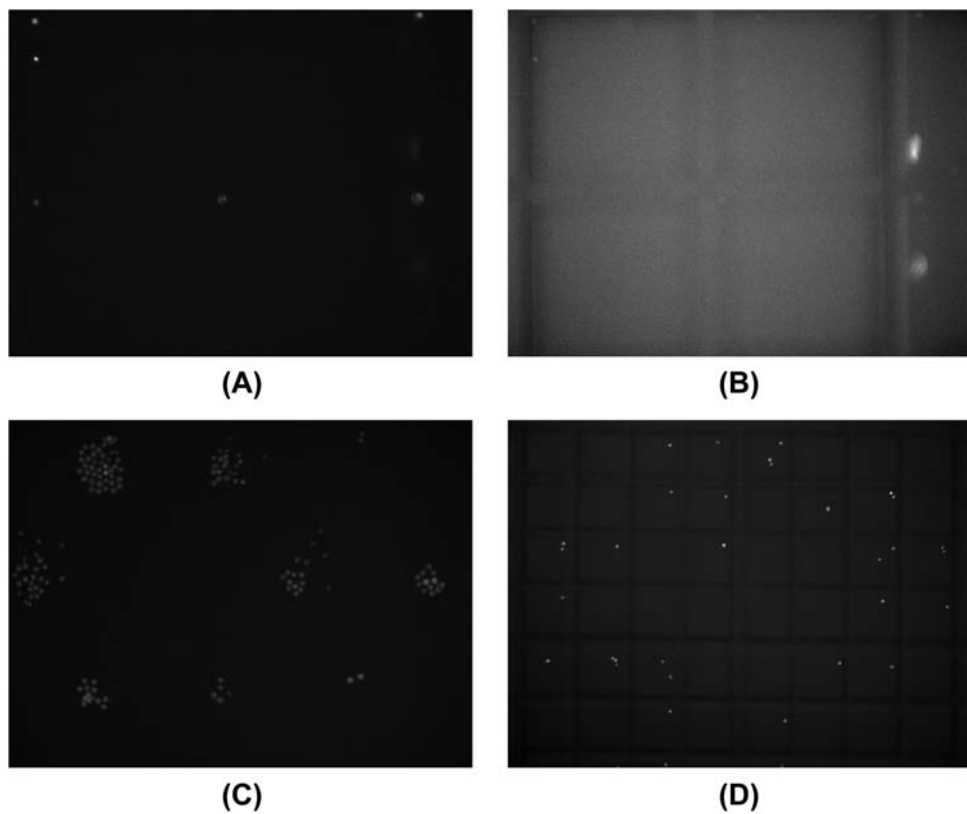
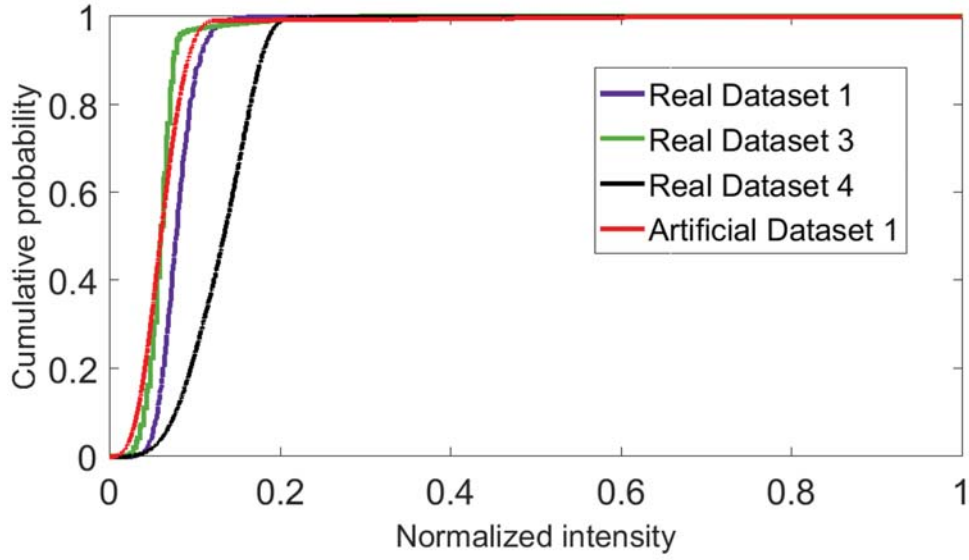
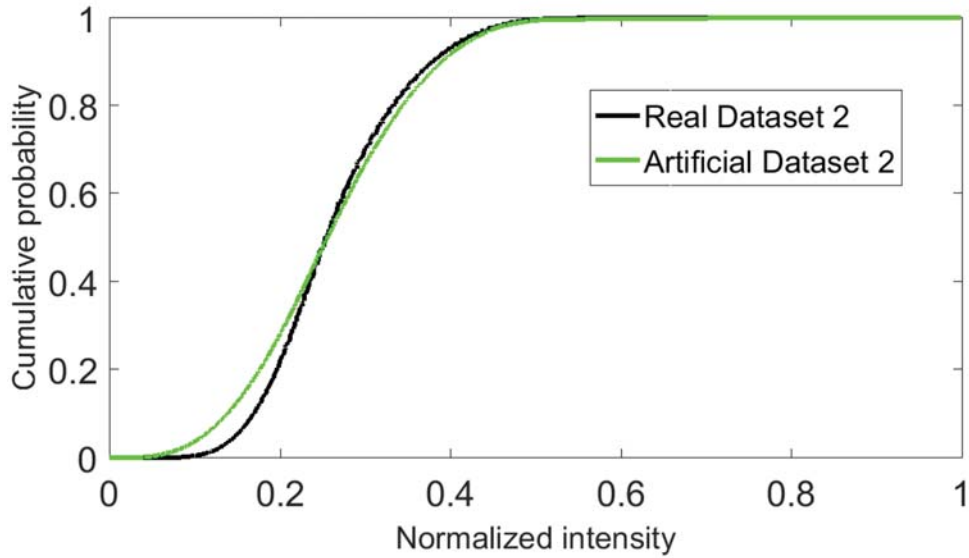


Figure 4: An image representative of real dataset 1 to 4 captured from: (a) rectangular well in FUCCI red channel, (b) rectangular well in FUCCI green channel, (c) hexagonal well in Ubiquitin-GFP channel, (d) square well in Ubiquitin-GFP channel.



(A)



(B)

Figure 5: Empirical cumulative probability distribution of the normalized intensity from real datasets and artificial datasets: (a) Artificial dataset 1 (red) (b) Artificial dataset 2 (green), indicating that they follow similar distributions.

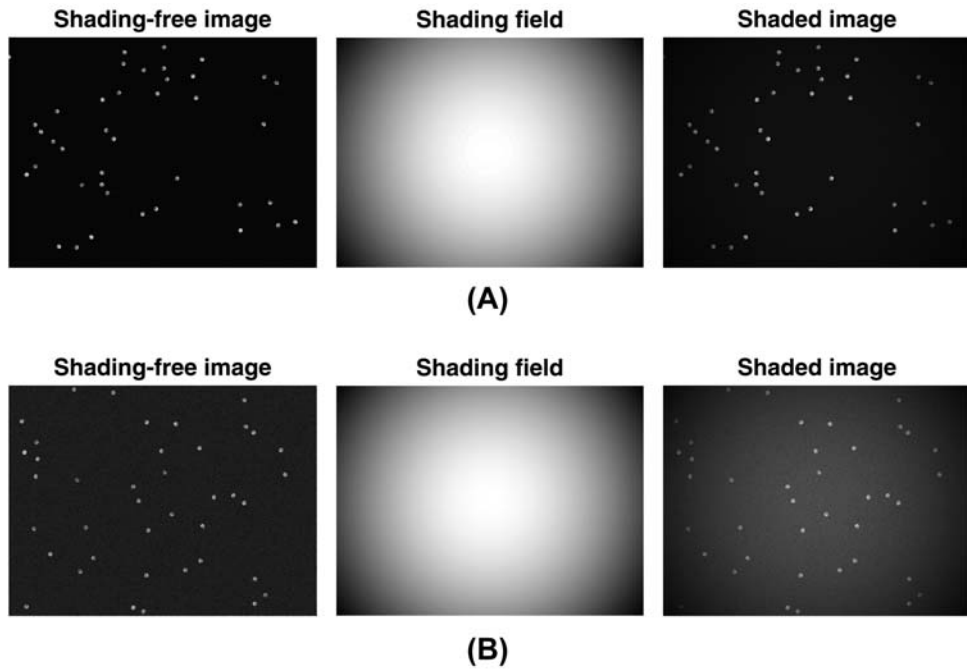


Figure 6: One shading-free image (left), corresponding shading field (middle) and resulting shaded image (right) from artificial dataset 1 (a) and artificial dataset 2 (b), respectively. Here the shaded image is approximately equal to the shading-free image multiplied by the shading field.

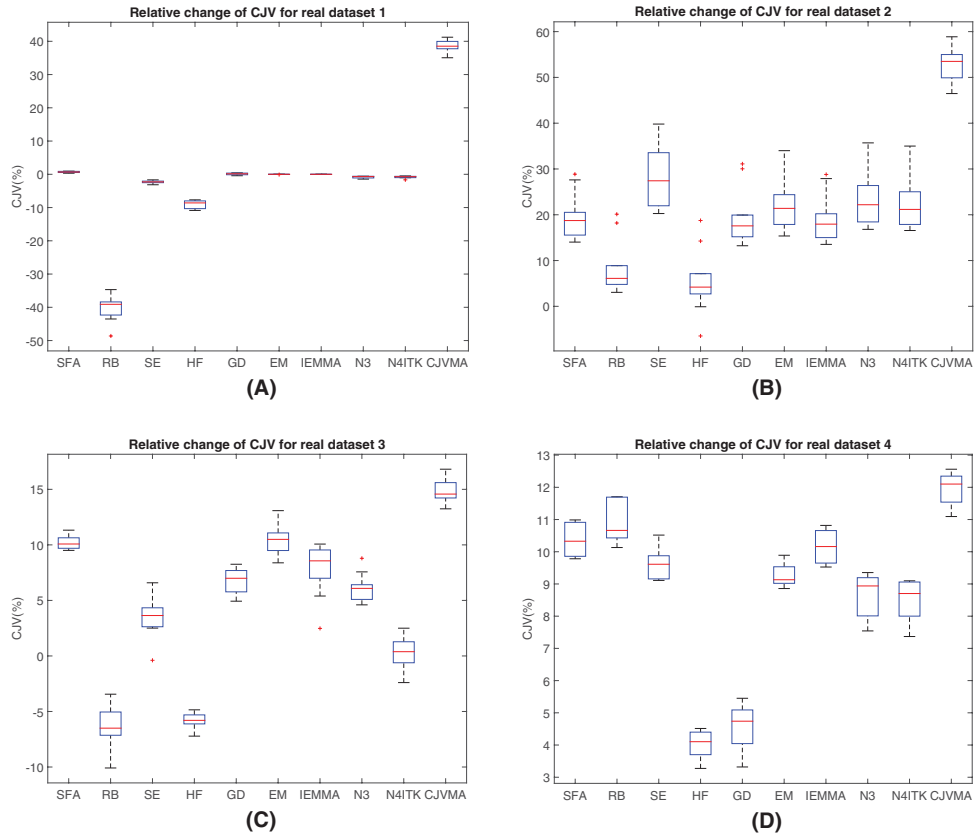


Figure 7: Distribution of changes of CJV for real fluorescence images corrected by ten shading correction methods. Each boxplot shows the relative change of CJV for ten real images. Larger value indicates higher performance of the method.

Table 1 Summary of different retrospective shading correction methods

Shading correction methods	Image formation model	Assumptions	Number of tuning parameters	Advantages	Disadvantages
Surface fitting (SF)	Multiplicative or Additive	The selected points could represent the global background	2	It performs well for images containing small-sized and medium-sized objects	The performance heavily relies on the selected points
Rolling ball (RB)	Multiplicative or Additive	The size of objects of interest is smaller than the background variation	1	It is simple implemented and relatively fast executed	It fails to correct biomedical images with large objects
Signal envelope (SE)	Additive	The shading component is slowly-varying and could be estimated from the signal envelope	2	It does not require the objects of interest to be small in size, and is widely used in biomedical images from microscopy to MR	It is used to remove additive shading component only
Homomorphic filtering (HF)	Multiplicative	The shading is low-frequency content which can be separated from the high-frequency spectrum of the true image	4	The shading field is not modelled by a parametric surface but estimated in the Fourier domain	It produces halo artefacts on the boundaries when the parameters of the filter are not tuned appropriately
Gradient distribution (GD)	Multiplicative	The image gradient probability distribution is sparse	5	Easy to implement and efficient for MR images	It is computational expensive and heavily dependent on the parameters of the weighting function
Entropy minimization (EM)	Multiplicative and Additive	Entropy increases after shading	0	No tuning parameters, and performs well for images with either large or small objects	High-order polynomial is required for accurate estimation of the shading field with a small local variation
Information minimization (IM)	Multiplicative and Additive	Entropy increases after shading	0	No tuning parameters and solely exploits the information naturally present in an image	It is optimization based and thus cannot generally guarantee the global optimum

Non-parametric non-uniform intensity normalization (N3)	Multiplicative	The true image intensities are independent identically distributed random variables	7	No a prior knowledge on the intensity probability distribution of the image structures and widely applied to almost any MR images	It may converge to local minima
N4ITK	Multiplicative	The true image intensities are independent identically distributed random variables	6	Improved convergence performance of the smooth shading field estimation compared to N3 and also widely applied to almost any MR images	It may converge to local minima and extremely computational expensive

Table 2 Summary of quantitative evaluation measures

Quantitative evaluations	Measure	Segmentation needed
Reference based evaluation	L_2 – norm	No
	Root mean squared error (RMSE)	No
	Spearman’s correlation coefficient (r)	No
	Peak signal-to-noise ratio (PSNR)	No
Reference-free based evaluation	Standard deviation (σ)	No
	Coefficient of variation (CV)	Yes
	Coefficient of joint variation (CJV)	Yes
	Signal-to-noise ratio (SNR)	No

Table 3 Implementation details of shading correction methods

Shading correction methods	Tuning parameters	Application tools	Alternative image model
Surface fitting (SFA)	$xOrder, yOrder$	“polynomial fit” in Fiji	Additive

Rolling ball (RB)	R	“Subtract Background” in Fiji	Additive
Signal envelope (SE)	M, δ	Matlab	Additive
Homomorphic filtering (HF)	$\gamma_H, \gamma_L, D_0, c$	“Homomorphic Filter” in Fiji	Multiplicative
Gradient distribution (GD)	$s, \alpha, D, M, \epsilon$	Matlab	Multiplicative
Entropy minimization (EM)	None	“Entropy Minimization” in Fiji	Multiplicative and Additive
Information minimization (IM)	None	“Entropy Minimization” in MIPAV	Multiplicative or Multiplicative and Additive
Non-parametric non-uniform intensity normalization (N3)	$\theta, M, F, s, \beta, \varphi, n$	“Inhomogeneity N3 Correction” in MIPAV	Multiplicative
N4ITK	$M, \beta, \varphi, n, r, N$	“Inhomogeneity N4 Correction” in MIPAV	Multiplicative

Table 4 Summary of correction methods implemented and parameter values used for artificial and real datasets

Shading correction methods	Tuning parameters	Two artificial datasets	Four real datasets
SFA	$xOrder, yOrder$	2, 2	2, 2
RB	R	18	18
SE	M, δ	28, 0.01	28, 0.01
HF	$\gamma_H, \gamma_L, D_0, c$	10, 0.1, 1, 80	10, 0.5, 1, 80
GD	$s, \alpha, D, M, \epsilon$	0.25, 0.1, 5, 4, 0.0001	0.25, 0.1, 5, 4, 0.0001
EM	None	—	—
IEMMA	None	—	—

N3	$\theta, M, F, s, \beta, \varphi, n$	Min+1, 100, 0.0001, Default, 2, 0.5, 0.01	Min+1, 100, 0.0001, Default, 2, 0.5, 0.01
N4ITK	$M, \beta, \varphi, n, r, N$	20, 0.0001, 0.5, 0.01, 2, 10	20, 0.0001, 0.15, 0.01, 3, 4

Table 5 Ranks of all shading correction methods based on different performance measures for artificial dataset 1. Rank “1” indicates the best performance among all methods, while “9” indicates the worst. This table shows that it is difficult to determine an optimal approach because the rank varies between performance measures.

Shading correction methods	$L_2 - norm$	RMSE	PSNR	SNR	CV_c	CV_b	$\overline{CV}(\%)$	σ_c	σ_b	$\overline{\sigma}(\%)$	CJV	CJV(%)	Number of wins
SFA	8	9	9	1	1	1	1	7	7	8	6	6	4
RB	9	1	1	9	9	9	9	8	6	7	7	7	2
SE	6	3	3	2	7	2	7	6	1	3	9	9	1
HF	7	8	8	8	8	8	8	9	9	9	8	8	0
GD	3	2	2	7	4	7	4	1	4	1	3	3	2
EM	5	5	5	6	6	5	6	4	2	4	5	5	0
IEMMA	4	6	6	5	5	4	5	5	5	6	4	4	0
N3	1	4	4	4	3	6	3	2	3	2	2	2	1
N4ITK	2	7	7	3	2	3	2	3	8	5	1	1	2

Table 6 Kendall's tau coefficient between L_2 – norm and other measures for artificial datasets

Data	L_2 – norm	RMSE	PSNR	SNR	CV_c	CV_b	$\overline{CV}(\%)$	σ_c	σ_b	$\overline{\sigma}(\%)$	CJV	CJV(%)
Dataset 1	1	0.056	0.056	0.167	0.556	0.056	0.556	0.722	0.111	0.444	0.667	0.667
Dataset 2	1	0.111	0.111	0.500	0.333	0.444	0.389	0.556	0.444	0.667	0.611	0.611
Mean Values	1	0.084	0.084	0.334	0.445	0.250	0.473	0.639	0.278	0.556	0.639	0.639

Table 7 Spearman's correlation between L_2 – norm and other measures for artificial datasets

Data	L_2 – norm	RMSE	PSNR	SNR	CV_c	CV_b	$\overline{CV}(\%)$	σ_c	σ_b	$\overline{\sigma}(\%)$	CJV	CJV(%)
Dataset 1	1	0.005	-0.005	-0.217	0.517	0.100	-0.517	0.883	0.250	-0.700	0.833	-0.833
Dataset 2	1	0.083	-0.083	-0.583	0.417	0.533	-0.517	0.683	0.583	-0.833	0.783	-0.783
Mean Values	1	0.044	0.044	-0.400	0.467	0.317	-0.517	0.783	0.417	-0.767	0.808	-0.808