Identification of Retinal Ganglion Cells from β -III Stained Fluorescent Microscopic Images

He Gai¹ · Yi Wang¹ · Leanne L.H. Chan¹ · Bernard Chiu¹

Published online: 23 July 2020

© Society for Imaging Informatics in Medicine 2020

Abstract

Optic nerve crush in mouse model is widely used for investigating the course following retinal ganglion cell (RGCs) injury. Manual cell counting from β -III tubulin stained microscopic images has been routinely performed to monitor RGCs after an optic nerve crush injury, but is time-consuming and prone to observer variability. This paper describes an automatic technique for RGC identification. We developed and validated (i) a sensitive cell candidate segmentation scheme and (ii) a classifier that removed false positives while retaining true positives. Two major contributions were made in cell candidate segmentation. First, a homomorphic filter was designed to adjust for the inhomogeneous illumination caused by uneven penetration of β -III tubulin antibody. Second, the optimal segmentation parameters for cell detection are highly image-specific. To address this issue, we introduced an offline-online parameter tuning approach. Offline tuning optimized model parameters based on training images and online tuning further optimized the parameters at the testing stage without needing access to the ground truth. In the cell identification stage, 31 geometric, statistical and textural features were extracted from each segmented cell candidate, which was subsequently classified as true or false positives by support vector machine. The homomorphic filter and the online parameter tuning approach together increased cell recall by 28%. The entire pipeline attained a recall, precision and coefficient of determination (r^2) of 85.3%, 97.1% and 0.994. The availability of the proposed pipeline will allow efficient, accurate and reproducible RGC quantification required for assessing the death/survival of RGCs in disease models.

Keywords Retinal ganglion cell (RGC) · Fluorescent microscopy · Cell segmentation · Homomorphic filter · Online parameter tuning

Introduction

The growth ability of axons in the mammalian central nervous system (CNS) is reduced after synaptic connections have been established shortly after birth [1]. The failure of axons to regenerate after an injury leads to permanent functional disabilities after the occurrence of spinal cord injury, brain trauma and the onset of multiple sclerosis. In the optical system, glaucoma is characterized by the damage of retinal ganglion cells (RGCs) and their axons due to intraocular pressure and is a leading cause of irreversible blindness worldwide [2]. The optic nerve crush injury mouse model has been widely used as an

experimental model for investigations relating to neuronal death and survival mechanisms in glaucoma, traumatic optic neuropathy and, more generally, axonal injuries in CNS [3, 4]. Many therapeutic approaches have been introduced to enhance the regenerative capability of RGC axons, such as gene deletion [5, 6], growth-promoting treatments [7, 8] and axonal guidance [9, 10]. Since the survival of RGC somata is a key indicator of positive effects of the treatments, the ability to identify and accurately count the RGC becomes essential in these investigations.

Retrograde fluorogold (FG) labelling and immunolabelling with an antibody directed against β -III tubulin are two commonly used techniques for visualizing and quantifying neuronal survival. Although the retrograde FG labelling is considered a specific and accurate method for RGC labelling, it is technically challenging as it requires an intracranial surgery to inject the FG tracer to the superior colliculi. Experiments involving the FG tracer requires a substantial incubation time as it takes 7 days for the tracer

✉ Bernard Chiu
bcychiu@cityu.edu.hk

¹ Department of Electrical Engineering, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong

to accumulate in the RGC soma [11]. β -III tubulin antibody is commonly used to label RGCs by highlighting the microtubules of the cytoskeleton [12]. In contrast to FG labelling, β -III tubulin staining is easier to be administered and has a shorter incubation time. The reliability of β -III tubulin in quantification for RGCs has been confirmed by a number of studies that co-localized β -III tubulin stained and FG-labelled RGCs [12, 13]. For these reasons, β -III tubulin is frequently used for identifying RGCs and for quantifying RGC loss in optic nerve injury models [14–17].

However, the need for quantifying RGCs in β -III tubulin stained images poses a challenge in image analysis. First, the quality of the images limits RGC quantification accuracy in manual and automated counting [18] as shown in the first row of Fig. 1. Inhomogeneous staining of RGCs occurs due to uneven tissue penetration [19] and the fact that the retinal structure may not be fully flattened onto

the imaging slide in experiments [18]. As a result, the cell boundary contrast is uneven and depends on the illumination level. In addition, structures such as the dendrites and axons are also highlighted in the immunostained images and partially block the view of a portion of RGCs. Second, as each image contains hundreds of cells, manual cell counting is time-consuming. Manual counting is also prone to observer variability.

Although many automated algorithms have been developed to segment digital pathology and microscopic images [20], Dordea et al. [18] introduced the only algorithm specifically designed to segment RGCs stained by a β -III tubulin antibody. This algorithm required training of a classifier using the *CellProfiler* interface [21]. Candidate regions were presented to a user, who then identified whether or not the regions represented actual cells. The GentleBoost algorithm was then applied to formulate a set

Fig. 1 An example image from each of the four mice before and after top-hat and homomorphic illuminations. Each column shows the original and the filtered versions of the same image. The magenta dots represent cell annotations made by an expert observer. The first row shows the original images. The second and third rows show the images filtered by the top-hat filter and the corresponding ridge-enhanced images, respectively. The fourth and fifth row show the images processed by the top-hat filter followed by the homomorphic filter and the corresponding ridge-enhanced images. The white contours shown in the second and fourth rows represent the candidate cells identified without and with homomorphic filtering respectively

of rules distinguishing positive and negative cells. Each rule thus formulated involved thresholding a cell feature. This training process iterated until the classification accuracy had stabilized. A limitation of this approach is that rules based on thresholding individual features did not take into account the interaction between different features and constrained the decision boundary to be perpendicular to axes in the high-dimensional feature space. While the algorithm produces a correlation with manual counting results ($r^2 = 0.64$) for images deemed “high-quality”, 28% of images were deemed “low-quality” and the correlation with manual counting results is much lower ($r^2 = 0.22$) for these “excluded” images. An edge detection method was applied to a binarized version of the original image. Images with a circularity score higher than an empirical threshold (0.13) were classified as having “high-quality”, whereas those below were classified as having “low-quality”. This criterion is expected to exclude many less illuminated images; for such images (e.g., the images at the first and second columns in Fig. 1), the algorithm would not generate accurate results.

The major focus of this work is to develop and validate an algorithm able to identify RGCs from β -III tubulin stained images with a wide range of illumination. Since the cell outline is often highlighted in images with cytoskeleton stained images, cell segmentation techniques can be developed out of existing tools capable of segmenting surface-stained cells. As described in a comprehensive review [20], cell segmentation pipeline typically consists of two stages: cell detection and cell segmentation. Typical cell detection techniques involve linear or non-linear spatial filtering, morphological operations and supervised methods, such as deep neural networks. After comparing the performance of 7 spatial filters, Hodneland et al. [22] concluded that the coherence enhancing filter [23] was more suitable than Gaussian filter, median filter, partial-differential-equation-based edge enhancing diffusion [24] and inverse diffusion [25] for highlighting ridges appearing along the boundaries of surface-stained cells. Therefore, the coherence enhancing filter was selected in the current study. Morphological operations [26] can be used in extracting boundaries from the filtered image with the edges highlighted, such as the ridge-enhanced images generated in this study. Deep neural networks have been used for constructing cell probability maps before cell segmentation [27, 28]. However, these techniques require a large number of manually annotated images for training and are not suitable for the current evaluation in which an expert took days to harvest and immunostain the retina and 30–40 minutes to annotate cells in each image. Pre-trained U-Net has now been shown to require fewer training samples in several data sets [29]. Although promising, the pre-trained model must be evaluated for its accuracy in the

detection of cells from retinal images with inhomogeneous staining across images and uneven illumination. For cell segmentation, the watershed and the level-set algorithms are the most widely used algorithms according to Tables II–V in [20]. These two algorithms have been compared in Hodneland et al. [22] with the watershed algorithm providing a slightly higher accuracy. For this reason, the watershed algorithm was applied in the current study.

Out of the existing cell segmentation software toolboxes, *CellSegm* [30] was selected in our preliminary attempt for segmenting RGCs from β -III tubulin stained images as it provides a pipeline for segmenting surface-stained cells, which involves state-of-the-art cell image processing and segmentation techniques including coherence enhancing diffusion, ridge enhancement and watershed segmentation. Segmented cells were classified to be actual and falsely detected cells based on thresholds for five geometric attributes of each cell. However, from our initial experience, we identified the following shortcomings of *CellSegm* in our application:

- It was not sensitive or precise in RGC detection, especially in less illuminated images.
- The accuracy in discriminating actual and falsely detected cells was not high. Furthermore, the thresholds used for discrimination were empirically selected, thereby limiting reproducibility. In addition, tuning these thresholds manually for each data set was time-consuming.
- The performance was highly dependent on the user-specified parameters involved in the framework.

This work introduces a framework for RGC identification that addresses these three issues. This tool will promote the efficiency, accuracy and reproducibility in future investigations involving the death and survival mechanisms of RGCs.

Methods

Figure 2 defines the workflow of the cell identification algorithm, which involves two stages: (i) the candidate segmentation stage and (ii) the cell identification stage, which are referred to as Stages I and II throughout the paper. For a high-performance cell segmentation framework, sufficient actual cells should be identified in Stage I and false positives should be removed in Stage II without discarding too many true positive cells. We added three modules to improve the cell segmentation and identification performance: (i) an illumination correction module to adjust for illumination difference across the image so as to enable the pipeline to segment more candidate cells, (ii) a framework to

Fig. 2 Flowchart of the proposed algorithm. Steps implemented in the *CellSegm* software are shaded in gray

remove false positives identified at Stage I. This false positive removal framework involves feature extraction, principal component analysis for feature dimensionality reduction

and support vector machine (SVM) for identifying actual cells from falsely detected cells and (iii) online and offline methods for tuning model parameters to increase cell discrimination accuracy. Model parameters include those listed in Table 1, detailed in “[CellSegm Components](#)”, and SVM parameters detailed in “[Feature Extraction, Dimensionality Reduction and Classification](#)”. These three new modules are described in the following sections. The improvement associated with each of the modules was evaluated, as described in “[Experimental Methods](#)”.

Subjects and Image Acquisition

Whole-mount mouse retinas were immunostained following standard protocols [16]. Rabbit anti-neuronal class β -III tubulin (TUBJ1, 1:800 dilution; Sigma-Aldrich) was used as the primary antibody. Alexa Fluor 488 (1:300 dilution, Molecular Probes) was applied as the secondary antibody. All antibodies were diluted in a solution consisting of 10% normal donkey serum (NDS) and 1% Triton X-100 in phosphate buffer solution (PBS). A fluorescent microscope (Eclipse 90i; Nikon) was used to randomly sample 8–12 fields per retina. This investigation involved 37 images acquired from four mice. Ethics approval for the study protocol was obtained from the Research Ethics Committee of City University of Hong Kong (Reference No. 2011SRG100). Figure 1 shows a representative image for each of the mice. To establish the range of inter-observer variability, two observers counted the stained RGCs and placed landmarks inside each detected cell using the interface available in ImageJ. The two observers were trained on retinal imaging and cell counting for two years and one year, respectively, in a laboratory focusing on axonal regeneration, molecular and cell biology investigations at City University of Hong Kong. The manual annotations produced by one of the observers were assigned to serve as the surrogate gold standard for training and evaluating the proposed algorithm, as described in “[Parameter Tuning](#)” and “[Experimental Methods](#)”. Each

Table 1 The set of 13 parameters involved in the candidate cell identification stage

Parameter	Description
D_{top}	Size of the circular structural element used in the top-hat filter
γ_H, γ_L	Maximum and minimum gain of the homomorphic filter
c, D_0	Control the sharpness of transition from γ_L to γ_H for the homomorphic filter
I_{max}	Number of iterations used in coherence enhancing diffusion (CED)
κ	Edge sensitivity in CED
R_D	The diameter of the Gaussian kernel used in ridge enhancement
R_{sd}	The standard deviation of the Gaussian kernel used in ridge enhancement
μ	A parameter used to determine the adaptive threshold defined in Eq. 4
γ	A parameter defining the size of the averaging filter used in adaptive thresholding (Eq. 4)
A_{max}, A_{min}	Maximum and minimum cell areas allowed in watershed segmentation

image had an intensity value from 0 to 255 and contained 104 to 271 cells according to this gold standard.

Proposed Illumination Correction

One of the major challenges of immunostaining is the inhomogeneous cell illumination resulting from uneven tissue penetration [19]. The first row of Fig. 1 shows four example images in which cells are unevenly illuminated. The uneven illumination would severely compromise the performance of the subsequent segmentation operations. Thus, the top-hat filter was introduced in *CellSegm* for illumination correction [26]:

$$\hat{I} = I - (I \circ S), \quad (1)$$

where $\hat{I}$ and I are the output and input images respectively, $\circ$ denotes the grayscale morphological opening operator and S is the circular structured element with a user-specified diameter D_{top} . The second row of Fig. 1 shows the output images from the top-hat filter. Although the background has been made darker and more even, a substantial number of cells with low illumination were merged into the background, making them difficult to be identified in subsequent segmentation operations. For this reason, we proposed to use a second filter known as the homomorphic filter to further enhance background homogeneity [26]. Homomorphic filtering reduces background inhomogeneity by applying the following high-pass filter in the log-domain:

$$H(u, v) = (\gamma_H - \gamma_L) \left[1 - \exp \left(-c \frac{u^2 + v^2}{D_0^2} \right) \right] + \gamma_L, \quad (2)$$

where u and v are the horizontal and vertical frequencies. The parameters γ_H , γ_L , c and D_0 represent the user-specified parameters of the homomorphic filter, which were later tuned to optimize cell identification performance. The gain of the filter was determined by γ_H and γ_L , whereas the sharpness of the transition from γ_H to γ_L was controlled by the parameters c and D_0 . The fourth row of Fig. 1 shows the example images after homomorphic filtering.

CellSegm Components

Three *CellSegm* modules were used in this paper, including the image smoothing, ridge enhancement and watershed segmentation modules. These techniques are described in detail elsewhere [22, 30] and briefly summarized here.

Coherence enhancing diffusion (CED) was used for smoothing. This is an iterative anisotropic diffusion algorithm that smooths along the edges of the cell but not across them [23]. Two user-specified parameters are involved in CED: the number of iterations, I_{max} , and the edge sensitivity, κ .

Ridge enhancement was performed by computing the eigenvalues of the convoluted Hessian matrix at each point [22]. The convoluted Hessian matrix was defined by:

$$H_g = g * \begin{bmatrix} (u_g)_{xx} & (u_g)_{xy} \\ (u_g)_{yx} & (u_g)_{yy} \end{bmatrix}, \quad (3)$$

where u_g is the Gaussian convoluted image and $g * ()$ is an additional Gaussian convolution. While there is a large intensity variation across the cell membrane, the variation is small along the tangential direction of the membrane. Pixels along the cell membrane can be characterized as ridges with $\lambda_1 < 0$ and $\lambda_2 \approx 0$, where λ_1 and λ_2 are the eigenvalues of the Hessian matrix ordered in descending absolute values [31]. An image $u_H(x) = -\lambda_1 - \lambda_2^2$ was generated to highlight ridges. The ridge-enhanced images generated for the images processed by the top-hat filter were plotted in the third row of Fig. 1, and those generated for images processed by the top-hat filter followed by the homomorphic filter were plotted in the fifth row of the figure. As the Hessian matrix was constructed by the derivatives of the Gaussian-smoothed image, the diameter and the standard deviation of the Gaussian kernel were required to be specified, which are denoted by R_D and R_{sd} .

Next, the marker-controlled watershed algorithm was applied to segment cells. Adaptive thresholding was first applied to binarize the ridge-enhanced image:

$$u_b(x) = \begin{cases} 1, & \text{if } u(x) > \mu \max(u) + A_\gamma(u, x) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where x represents the coordinates of a pixel in an image u , $\max(u)$ is the maximum image intensity in the image u and $A_\gamma(u, x)$ is the average image intensity in a γ -neighbourhood around x . μ and γ are user-defined parameters. This thresholding approach is adaptive because the threshold used depends on the position of a pixel. Disconnected cell boundaries in u_b were closed and the cells were filled by iterative morphological closing and filling operations performed using circular structural elements with increasing radii. These operations produced a set of “markers” or potential cell regions [22]. Markers larger than the user-specified maximum area threshold A_{max} , or smaller than the user-specified minimum area threshold A_{min} were removed. Initialized by these markers, the watershed segmentation algorithm was applied to extract the cell boundaries.

Feature Extraction, Dimensionality Reduction and Classification

The existing *CellSegm* software classified candidate regions into cell and background regions by thresholding six features, including the maximum and minimum areas of the cell, mean image intensities inside the cell and along

the boundary, convex area and convex perimeters [22]. The use of thresholding on individual features did not take into account the interaction between different features and constrained the decision boundaries to be perpendicular to the axes in the multi-dimensional feature space, leading to suboptimal cell identification performance. In this paper, we used a combination of PCA and SVM based on 31 geometric, statistical and textural features extracted for each cell as described below. PCA was used to reduce the dimensionality of the feature space before SVM was applied. Beside dimensionality reduction, PCA also reduces redundancy by minimizing the covariances of different features in the reduced space. In our implementation, the manual cell annotations on a randomly selected image from each mouse were used to train the SVM with a radial basis kernel: $K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$, which involved a user-specified parameter σ that determined the width of the kernel. Another user-specified parameter was the penalty parameter C , which determined how much penalty that the cost function assigned to misclassified points [32]. For large values of C , the cost function of SVM is optimized by a boundary that minimizes misclassification of training points and has a smaller margin between the actual and falsely detected cells in the feature space. For small values of C , the cost function is optimized by a boundary with a larger margin between the actual and falsely detected cells, even though the boundary misclassifies more training points. For these reasons, C is a parameter controlling the overfitting of the training data. These two parameters were tuned later to optimize identification performance.

The 31 features extracted in this work consist of 10 geometric features, 6 statistical features and 15 textural features. The geometric features include area, perimeter, major and minor axis lengths, convex area, eccentricity and solidity computed by the `regionprop` function in Matlab (Natick, MA). The remaining three geometric features, including circularity, rectangularity and elongation, are defined as follows:

$$\text{Circularity} = \frac{4\pi A}{P^2} \quad (5)$$

$$\text{Rectangularity} = \frac{A}{L_1 L_2} \quad (6)$$

$$\text{Elongation} = \frac{\min\{L_1, L_2\}}{\max\{L_1, L_2\}} \quad (7)$$

where A , P , L_1 and L_2 are the area, perimeter, the major and minor axis lengths of the cell respectively.

Statistical features describe the properties of the intensity histogram constructed for the region inside each cell candidate. The statistical features computed in this work include the mean, variance, smoothness, third moment, uniformity and entropy defined in Table 11.2 of [33].

The textural features are computed based on the gray-level gradient co-occurrence matrix (GGCM) proposed in Ubul et al. [34]. Instead of only considering the gray-level distribution as in the gray-level co-occurrence matrix, GGCM considers the co-occurrence of specific gray-levels and gradient magnitudes. GGCM parameters were derived based on the two-dimensional histogram $h(L, Grad)$, where L denotes the gray level and $Grad$ denotes the gradient magnitude. The following 15 parameters derived from h were extracted for each cell: small gradient dominance, large gradient dominance, gray asymmetry, gradient asymmetry, energy, gray-level mean, gradient mean, gray-level variance, gradient variance, correlation, gray-level entropy, gradient entropy, mixed entropy, inertia and homogeneity, as defined in Table 1 of [34].

Parameter Tuning

Offline Parameter Tuning Based on a Training Image

Figure 3 shows a flowchart illustrating procedures involved in online and offline tuning. The following 14 parameters were required to be tuned: 13 parameters listed in Table 1, which are collectively represented by P in Fig. 3, and 2 SVM parameters: σ and C . The left panel of Fig. 3 illustrates how this set of parameters was tuned in the training stage. This tuning process is referred to as offline tuning hereafter. One image from each of the four mice was involved in offline parameter tuning and was excluded from the evaluation. These parameters were initialized empirically and optimized sequentially by changing a single parameter while keeping the remaining parameters constant. The recall and the precision corresponding to each set of parameters were obtained by comparing the algorithm segmentation and the manual annotation of the cell centroids. False positives (FP) were identified when a segmented contour did not enclose a manually identified cell centroid. False negatives (FN) were cases where a manually identified cell centroid was not enclosed by a segmented contour. True positives (TP) were observed when a segmented contour enclosed a manually identified cell centroid. In our experiment, we did not have cases where multiple centroids were enclosed by a single segmented contour (examples are shown in Fig. 1). This is attributable to the superior performance of the ridge detection and segmentation algorithm built-in to *CellSegm*. The parameter set corresponding to the maximum F_1 score was declared as optimum. The F_1 score is defined as:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (8)$$

The optimal parameter set is denoted by $\hat{P}$, $\hat{C}$, and $\hat{\sigma}$ in Fig. 3. Since this is a sequential optimum search, the

Fig. 3 Flowchart illustrating procedures involved in online and offline tuning. P denotes the set of 13 parameters involved in Stage I and listed in Table 1. ΔP is the change applied in online tuning. C_{MAN} denotes the set of cells identified manually. $C_{SVM}(P, C, \sigma)$ denotes the set of cells selected by SVM, which is a function of P , C , the penalty parameter in SVM, and σ , the standard deviation of the SVM kernel. $C_{KNN}(P, K)$ denotes the set of cells selected by KNN, which is a function of P and K , the number of neighbours involved. Recall and precision are denoted by Rec and $Prec$. The hat notation ($\hat{\cdot}$) represents the optimized version of the corresponding parameter. The optimal parameters for KNN, SVM and P were obtained in the training stage (offline tuning), and then P was fine-tuned by online tuning

optimized parameter set is not guaranteed to be the global optimum.

Online Fine-tuning of Parameters for Each Test Image

Since the inhomogeneity in cell illumination and the cell size vary significantly across images, even for images acquired from the same mouse, optimal recall and precision could not be attained with the parameters obtained by offline tuning. To address this issue, we developed a method to fine-tune the parameters involved in generating the candidate cells (i.e., P in Fig. 3) when segmenting each individual image, which we refer to as online tuning. Inspired by ensemble learning [35], this parameter tuning method involved two classifiers: the SVM and the K-nearest-neighbour (KNN) classifiers. A “better” parameter set P would give rise to an image with higher quality from which cells are easier to be detected, or in other words, the image would consist of more “obvious” cell candidates. Although SVM and KNN may generate different classification errors,

they would agree on classifying these “obvious” candidates to be cells. Thus, the degree of concurrence between the SVM and KNN classifiers can serve as a metric to quantify image quality, which can be optimized for each test image in the absence of the ground truth during the testing stage.

The right panel of Fig. 3 shows the steps involved in the online parameter tuning procedure. Similar to how SVM parameters and P were tuned, the parameter K of the KNN classifier was tuned in the training stage, resulting in an optimal K denoted by $\hat{K}$. Online tuning of P was implemented in a similar way as offline tuning, except that the F_1 score was obtained by comparing the SVM and KNN results as illustrated at the bottom right-hand corner of Fig. 3. Note that which segmentation was treated as “truth” has no effect on the F_1 score. Here, the F_1 score quantified the degree of concurrence between the SVM and KNN classifiers, and was maximized in the sequential optimization process.

Experimental Methods

Three modules were introduced in this paper to improve cell segmentation performance: a new illumination correction module, a new cell identification module (Stage II) and the parameter tuning strategy. Experiments were designed to quantify the improvement associated with each of these three modules. The recall and precision associated with the cell segmentation results generated in all settings described below were obtained by comparisons with manually identified cell centroids with TP, FP and FN defined in “Offline Parameter Tuning Based on a Training Image”. The coefficient of determination (r^2) was computed by squaring the correlation coefficient between the surrogate ground truth of the cell count and that obtained in one of the experimental settings. As the framework consists of two stages as shown in Fig. 2, the 7 experimental settings described below are denoted in the form (Stage I Option(s), Stage II Option(s)), with available Stages I and II options described below. Tukey’s test was used for pairwise comparisons among these 7 experimental settings. This test is essentially a t-test, except that it corrects for the family-wise type I error rate [36].

The effect of the illumination correction module was assessed by comparing the following two experimental settings. In both settings, smoothing, ridge enhancement and watershed segmentation were performed. Only the top-hat filter was used for cell illumination in the first setting, whereas the output of the top-hat filter was further processed by the homomorphic filter in the second setting. The parameters involved in the illumination, smoothing, ridge enhancement and watershed segmentation algorithms were tuned with a training image (i.e., by offline tuning). Since the purpose of this comparison was to assess the effect

of the homomorphic filter, Stage II and online parameter tuning were not performed here. The first setting is denoted by $\langle \text{TH}, \emptyset \rangle$ and the second is denoted by $\langle \text{TH} + \text{Homo}, \emptyset \rangle$, where TH and Homo represent top-hat and homomorphic filtering respectively. The fact that Stage II processing was not carried out is explicitly indicated by the null symbol.

The effect of the new cell classifier (Stage II) was evaluated by comparing the following three experiment settings. These three settings were built on the previous setting involving both top-hat and homomorphic filters. In the setting denoted by $\langle \text{TH} + \text{Homo}, \text{CellSegm} \rangle$, the cell classification algorithm built-in to *CellSegm* was used, whereas in the setting denoted by $\langle \text{TH} + \text{Homo}, \text{SVM} \rangle$, the SVM classification algorithm was used. The six user-specified parameters used in *CellSegm*, including the maximum and minimum area of the cell, mean image intensities inside the cell and along the boundary, convex area and convex perimeters defined in [22], were all tuned offline. In the setting denoted by $\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$, PCA was performed to reduce the dimensionality of the feature vector before SVM was applied. The low-dimensional representation generated by PCA retained 90% of the variances. The effect of PCA was quantified by comparing the results generated in the settings $\langle \text{TH} + \text{Homo}, \text{SVM} \rangle$ and $\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$. Online parameter tuning was not performed in these three settings.

The effect of the online parameter tuning strategy was assessed by first comparing the settings $\langle \text{TH} + \text{Homo} + \text{Online}, \emptyset \rangle$ and $\langle \text{TH} + \text{Homo}, \emptyset \rangle$ and then by the comparing settings $\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$ and $\langle \text{TH} + \text{Homo} + \text{Online}, \text{PCA} + \text{SVM} \rangle$. The setting $\langle \text{TH} + \text{Homo} + \text{Online}, \text{PCA} + \text{SVM} \rangle$ was built on $\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$ but involved online parameter tuning.

Observer cell annotation is associated with inter-rater variability. Segmentation of the algorithm should be interpreted with this variability taken into account. To establish the range of inter-rater variability, the image sets of three mice were annotated by a second observer and his annotation was compared with the surrogate gold standard generated by the first observer. The recall and precision

obtained in this comparison can serve as a benchmark when interpreting the results generated by the proposed framework.

Results

Table 2 shows the recall, precision and r^2 obtained in all experimental settings. Figure 1 shows an example image from each of the four mice with manually annotated cells represented by magenta. Candidate cells segmented without homomorphic filtering are shown as white contours in the second row, whereas those segmented with homomorphic filtering are shown in the third row. A qualitative comparison between the second and the third rows show that, when homomorphic filtering was not applied (i.e., $\langle \text{TH}, \emptyset \rangle$), many cells in the less illuminated regions were missed, whereas homomorphic filtering brightens darker regions, thereby making cells in these regions more readily identified. Tukey's comparison test shows that the recall and precision associated with $\langle \text{TH} + \text{Homo}, \emptyset \rangle$ were significantly larger than those obtained in $\langle \text{TH}, \emptyset \rangle$ (p values for recall and precision were both $< 10^{-12}$).

Since the parameters used in the illumination correction and the segmentation procedures were the same in the four settings involving Stage I options TH + Homo, the cell candidates segmented in these settings were the same. With different classifiers applied in three of these four settings, the sensitivities in these three settings were inevitably lower than the setting $\langle \text{TH} + \text{Homo}, \emptyset \rangle$ as Stage II aimed at reducing false positive detection.

Figure 4 shows the results obtained in Stage II for the three settings, with red, blue and yellow dots representing TP, FP and FN generated in Stage II. Recall is lower if more FNs are generated in Stage II (i.e., yellow dots in Fig. 4). The number of FNs is higher in $\langle \text{TH} + \text{Homo}, \text{CellSegm} \rangle$ than in $\langle \text{TH} + \text{Homo}, \text{SVM} \rangle$ and $\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$. Based on Tukey's multiple comparison tests, the reduction in recall as compared with $\langle \text{TH} + \text{Homo}, \emptyset \rangle$ was statistically significant for $\langle \text{TH} + \text{Homo}, \text{CellSegm} \rangle$

Table 2 Recall, precision in percentage and coefficient of determination (r^2) between manual and algorithm cell counts attained in different experimental settings

Settings	Recall	Precision	r^2
$\langle \text{TH}, \emptyset \rangle$	59.4 ± 4.86	61.3 ± 4.34	0.958
$\langle \text{TH} + \text{Homo}, \emptyset \rangle$	75.7 ± 4.40	70.2 ± 3.08	0.973
$\langle \text{TH} + \text{Homo}, \text{CellSegm} \rangle$	56.2 ± 3.03	81.9 ± 4.06	0.968
$\langle \text{TH} + \text{Homo}, \text{SVM} \rangle$	71.7 ± 4.12	93.6 ± 2.27	0.977
$\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$	74.5 ± 4.33	97.0 ± 1.29	0.978
$\langle \text{TH} + \text{Homo} + \text{Online}, \emptyset \rangle$	87.5 ± 2.63	70.6 ± 4.07	0.972
$\langle \text{TH} + \text{Homo} + \text{Online}, \text{PCA} + \text{SVM} \rangle$	85.3 ± 2.29	97.1 ± 1.29	0.994
Observer 2	94.1 ± 3.17	82.7 ± 5.20	0.956

Fig. 4 An example image from each of the four mice showing cell identification results. Each column shows the top-hat and homomorphic filtered images with annotations and contours superimposed. This figure shows the same images as in Fig. 1 and the order of the column in these two images correspond. The white contours represent all the cell candidates identified at Stage I. Since the first three rows involve the same cell candidate segmentation technique, cells segmented in these three settings are the same. The red, blue and yellow dots represent true positives, false positives and false negatives generated by Stage II, respectively. White contours with no dot represent true negatives. Arrows in the last row point to cells that were missed by the algorithm

($p < 10^{-12}$) and $\langle \text{TH} + \text{Homo}, \text{SVM} \rangle$ ($p = 0.004$), but not for $\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$ ($p = 0.93$). This results indicate that true positives identified in Stage I were mostly retained by $\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$, whereas the classification technique built-in to *CellSegm* discarded a substantial amount of annotated cells.

Conversely, the precision would be improved substantially if the FPs detected in Stage I are properly removed in Stage II. The precision is lower if more FPs remains after processed by Stage II (i.e., blue dots in Fig. 4). A qualitative comparison shows that much more blue dots remained in $\langle \text{TH} + \text{Homo}, \text{CellSegm} \rangle$ than in $\langle \text{TH} + \text{Homo}, \text{SVM} \rangle$ and $\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$. Tukey's multiple comparison tests shows that the increases in precision for all three settings involving Stage II were statistically significant as compared with $\langle \text{TH} + \text{Homo}, \emptyset \rangle$ (all p values $< 10^{-12}$). The precision for $\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$ was 97%. $\langle \text{TH} + \text{Homo}, \text{CellSegm} \rangle$ had a significantly lower precision than $\langle \text{TH} + \text{Homo}, \text{SVM} \rangle$ and $\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$ (both p values $< 10^{-12}$) and the precision of $\langle \text{TH} + \text{Homo}, \text{PCA} + \text{SVM} \rangle$ was significantly higher than $\langle \text{TH} + \text{Homo}, \text{SVM} \rangle$ ($p = 0.0007$).

With online parameter tuning applied, the average recall obtained in Stage I was substantially improved from 75.7% in $\langle \text{TH} + \text{Homo}, \emptyset \rangle$ to 87.5% in $\langle \text{TH} + \text{Homo} + \text{Online}, \emptyset \rangle$. A qualitative comparison between the third and the fourth rows of Fig. 4 shows that with the use of online parameter

tuning, more cells were identified in the less illuminated regions in Stage I. With the use of the PCA and SVM pipeline in the setting $\langle \text{TH} + \text{Homo} + \text{Online}, \text{PCA} + \text{SVM} \rangle$, the average recall decreased slightly to 85.3%, which was not statistically different from $\langle \text{TH} + \text{Homo} + \text{Online}, \emptyset \rangle$ ($p = 0.29$) based on Tukey's multiple comparison tests. The improvements in precision for $\langle \text{TH} + \text{Homo} + \text{Online}, \text{PCA} + \text{SVM} \rangle$ was significant as compared with $\langle \text{TH} + \text{Homo} + \text{Online}, \emptyset \rangle$ ($p < 10^{-12}$).

Compared with the gold standard annotation made by Observer 1, Observer 2 achieved an average recall of 94.1% and an average precision of 82.7%. The recall of the proposed algorithm (Setting $\langle \text{TH} + \text{Homo} + \text{Online}, \text{PCA} + \text{SVM} \rangle$) was $\approx 10\%$ lower than that of the human observer. Many of the cells missed by the proposed algorithm but not the human observers were predominantly those lying on the border of the images; these cells appear as a semi- or incomplete ellipsoid and failed to be captured by our cell identification stage as the pipeline was designed to identify closed contours (blue arrows in the last row of Fig. 4). A few cells with the whole cell body stained were not identified in Stage I as *CellSegm* was designed to detect regions with ridges enhanced (red arrows in the last row of Fig. 4), but these cells were rarely seen. A few cells with incomplete boundaries and poor illumination were missed (brown arrows the last row of Fig. 4). The precision generated by the algorithm was significantly higher than

Observer 2 ($p < 10^{-12}$). r^2 associated with the proposed algorithm was substantially higher than Observer 2 (0.994 vs. 0.956), indicating that the performance of the algorithm is well within the range of inter-observer variability.

Discussion

The motivation of this work stems from the need for an automated method to quantify RGCs from β -III tubulin stained images efficiently and reproducibly, which is important for the evaluation of therapeutic approaches aiming at enhancing RGC regeneration after injuries. Although cell segmentation techniques are available as open-source software [18, 30], the existing software does not provide high recall and precision in RGC recognition. This paper introduces an RGC segmentation and identification pipeline, which includes three innovative features: (i) a homomorphic filter highlighting the edges of the cell by removing low-frequency components attributable to uneven illumination, (ii) a module that extracted 31 cell features, reduced the dimensionality of cell feature vectors using PCA before RGC identification using SVM, (iii) online and offline parameter tuning methods for improving the accuracy of cell identification. The homomorphic filter and the online parameter tuning method together boosted RGC recall from 59.4 to 87.5%. We observed that the cell segmentation accuracy was sensitive to the model parameters; the image processing parameters (i.e., $\hat{P}$ in Fig. 3) giving rise to optimal segmentation results were highly image-specific, even for images acquired from the same mouse. For this reason, parameters tuned offline (i.e., in the training stage) were suboptimal for segmenting individual images. A highlight of the proposed pipeline is the introduction of an online process that tuned the model parameters in the absence of ground truth. While SVM and KNN made different errors, we hypothesized that they would concur more if the image processing parameters were optimized. For this reason, the degree of concurrence between the two classifiers was used as an objective function in online tuning. This technique contributed to an 11% increase of the RGC detection recall. Apart from being sensitive, a high-performance cell identification pipeline should also be precise. The experimental results show that the precision improved by extracting textural as well as morphological cell features and reducing the dimensionality of the feature vector. Implementation of PCA and SVM in Stage II increased the precision to 97.0% and 97.1% in the (TH + Homo, PCA + SVM) and (TH + Homo + Online, PCA + SVM), respectively.

As cell count is used as the primary parameter to quantify cell viability after optic nerve crush injury, the coefficients of determination (r^2) between manual and algorithm cell

counts were computed for various experimental settings. We also computed r^2 between cell counts obtained by Observers 1 and 2 to quantify the performance of the proposed algorithm in terms of inter-observer variability. r^2 obtained between annotations generated by the two observers was substantially lower than the r^2 between the annotation generated by the proposed algorithm and Observer 1. Since the algorithm was trained by Observer 1's annotations, and therefore, subject to the biases of Observer 1, the algorithm is not necessarily "better" than another human observer. However, our algorithm produced a result that fell within the range of observer variability. In longitudinal studies of cell viability after optic nerve crush injury, bias is less of a problem in monitoring changes as long as the cell counts from baseline and the follow-up images are reproducible. The higher efficiency and reproducibility afforded by the proposed algorithm makes it more suitable than manual annotations by different observers in quantifying RGC viability in a longitudinal study.

A limitation of this study is related to the use of a relatively small number of images in this pilot evaluation study of a novel algorithm. In this study, an expert experimenter took two days to harvest, process and immunostain the retina. Each observer took 30–40 min to count the RGCs under the fluorescent microscope and we require two observers to count the RGCs in each image to allow an assessment of inter-observer variability. For these reasons, manual processing for 37 images already took a considerable amount of time. While we acknowledge that a future study should be performed to assess the performance of the algorithm for more images with a wider range of image quality and with different stain and acquisition protocols, it should be noted that cell identification was performed on a cell-by-cell basis in this investigation. Each image contains 150 cells on average and more than 4000 cells were involved in the validation of the proposed framework. This data set has allowed validation of the proposed framework in a cell population with a broad range of geometric and textural variations.

Conclusions

A highly sensitive and precise tool was developed for identifying RGC from β -III tubulin stained retinal images. Technical innovations include the introduction of the homomorphic filter and an online segmentation parameter tuning model in the cell identification stage, which together increased the cell detection recall by close to 30%, and a classifier that removed most of the false positives and attained a precision of 97%. This tool will allow accurate and reproducible RGC quantification necessary

for assessing the death or survival of RGCs in disease models and evaluating regenerative therapies. In addition, the improvement over *CellSegm* demonstrated in this study suggests that the pipeline will benefit cell recognition performance in surface-stained fluoroscopic imaging, which is widely used because, in contrast to nuclear staining, surface-stained imaging allows for the detection of whole cells in closely packed arrangements and allows cell-level characterization of tissue and disease developments.

Acknowledgements The authors would like to acknowledge Dr. Eddie Ma and Mr. Bennett Au from Department of Biomedical Sciences at City University of Hong Kong for their help in establishing the TUJ1 immunostaining protocol.

Funding Information Dr. Chiu received funding support from the Research Grant Council of the HKSAR, China (Project No. CityU 11205917), and the City University of Hong Kong Strategic Research Grants (No. 7005226).

Compliance with Ethical Standards Ethics approval for the study protocol was obtained from the Research Ethics Committee of City University of Hong Kong (Reference No. 2011SRG100).

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Curcio M, Bradke F: Axon regeneration in the central nervous system: Facing the challenges from the inside. *Annu Rev Cell Dev Biol* 34:495–521, 2018
- Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY: Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology* 121:2081–2090, 2014
- Pernet V, Schwab ME: Lost in the jungle: new hurdles for optic nerve axon regeneration. *Trends Neurosci* 37:381–387, 2014
- Tuszynski MH, Steward O: Concepts and methods for the study of axonal regeneration in the CNS. *Neuron* 74:777–791, 2012
- Kurimoto T, Yin Y, Omura K, Gilbert HY, Kim D, Cen LP, Moko L, Kügler S., Benowitz LI: Long-distance axon regeneration in the mature optic nerve: contributions of oncomodulin, camp, and pten gene deletion. *J Neurosci* 30:15654–15663, 2010
- Sun F, Park KK, Belin S, Wang D, Lu T, Chen G, Zhang K, Yeung C, Feng G, Yankner BA, He Z: Sustained axon regeneration induced by co-deletion of PTEN and SOCS3. *Nature* 480:372–375, 2011
- Goldberg JL, Espinosa JS, Xu Y, Davidson N, Kovacs GTA, Barres BA: Retinal ganglion cells do not extend axons by default: promotion by neurotrophic signaling and electrical activity. *Neuron* 33:689–702, 2002
- Liu K, Tedeschi A, Park KK, He Z: Neuronal intrinsic mechanisms of axon regeneration. *Annu Rev Neurosci* 34:131–152, 2011
- Ersikine L, Herrera E: The retinal ganglion cell axon's journey: insights into molecular mechanisms of axon guidance. *Dev Biol* 308:1–14, 2007
- Pernet V, Joly S, Dalkara D, Jordi N, Schwarz O, Christ F, Schaffer DV, Flannery JG, Schwab ME: Long-distance axonal regeneration induced by cntf gene transfer is impaired by axonal misguidance in the injured adult optic nerve. *Neurobiol Dis* 51:202–213, 2013
- Huang TL, Huang SP, Chang CH, Lin KH, Sheu MM, Tsai RK: Factors influencing the retrograde labeling of retinal ganglion cells with fluorogold in an animal optic nerve crush model. *Ophthalmic Res* 51:173–178, 2014
- Jiang SM, Zeng LP, Zeng JH, Tang L, Chen XM, Wei X: β -III-Tubulin: a reliable marker for retinal ganglion cell labeling in experimental models of glaucoma. *Int J Ophthalmol* 8:643–652, 2015
- Chen H, Wei X, Cho KS, Chen G, Sappington R, Calkins DJ, Chen DF: Optic neuropathy due to microbead-induced elevated intraocular pressure in the mouse. *Invest Ophthalmol Vis Sci* 52:36–44, 2011
- Fitzgerald M, Bartlett CA, Evill L, Rodger J, Harvey AR, Dunlop SA: Secondary degeneration of the optic nerve following partial transection: the benefits of lomerizine. *Exp Neurol* 216:219–230, 2009
- Hu Y, Cui Q, Harvey AR: Interactive effects of c3, cyclic amp and ciliary neurotrophic factor on adult retinal ganglion cell survival and axonal regeneration. *Mol Cell Neurosci* 34:88–98, 2007
- Smith PD, Sun F, Park KK, Cai B, Wang C, Kuwako K, Martinez-Carrasco I, Connolly L, He Z: SOCS3 deletion promotes optic nerve regeneration in vivo. *Neuron* 64:617–623, 2009
- Pimentel B, Sanz C, Varela-Nieto I, Rapp UR, De Pablo F, de La Rosa EJ: c-Raf regulates cell survival and retinal ganglion cell morphogenesis during neurogenesis. *J Neurosci* 20:3254–3262, 2000
- Dordea AC, Bray MA, Allen K, Logan DJ, Fei F, Malhotra R, Gregory MS, Carpenter AE, Buys ES: An open-source computational tool to automatically quantify immunolabeled retinal ganglion cells. *Exp Eye Res* 147:50–56, 2016
- Li W, Germain RN, Gerner MY: Multiplex, quantitative cellular analysis in large tissue volumes with clearing-enhanced 3D microscopy. *Proc Natl Acad Sci USA* 114:E7321–E7330, 2017
- Xing F, Yang L: Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review. *IEEE Rev Biomed Eng* 9:234–263, 2016
- Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, Castoreno AB, Eggert US, Root DE, Golland P, Sabatini DM: Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. *Proc Natl Acad Sci USA* 106:1826–1831, 2009
- Hodneland E, Bukoreshtliev NV, Eichler TW, Tai XC, Gurke S, Lundervold A, Gerdes HH: A unified framework for automated 3-D segmentation of surface-stained living cells and a comprehensive segmentation evaluation. *IEEE Trans Med Imaging* 28:720–738, 2009
- Weickert J: Coherence-enhancing diffusion filtering. *Int J Comput Vis* 31(2):111–127, 1999
- Weickert J: A review of nonlinear diffusion filtering. In: ter Haar Romeny B, Florack L, Koenderink J, Viergever M Eds. *Scale-Space Theory in Computer Vision*. Springer, Berlin, 1997, pp 1–28
- Kimmel R, Malladi R, Sochen N: Images as embedded maps and minimal surfaces: movies, color, texture, and volumetric medical images. *Int J Comput Vis* 39(2):111–129, 2000
- Gonzalez RC, Woods RE: Digital image processing, 2nd edition. Englewood Cliffs: Prentice-Hall, 2002
- Xing F, Xie Y, Yang L: An automatic learning-based framework for robust nucleus segmentation. *IEEE Trans Med Imaging* 35:550–566, 2016
- Al-Kofahi Y, Zaltsman A, Graves R, Marshall W, Rusu M: A deep learning-based algorithm for 2-D cell segmentation in microscopy images. *BMC Bioinform* 19:365, 2018

29. Falk T, Mai D, Bensch R, Çiçek z., Abdulkadir A, Marrakchi Y, Böhm A, Deubner J, Jäkel Z, Seiwald K, Dovzhenko A, Tietz O, Dal Bosco C, Walsh S, Saltukoglu D, Tay TL, Prinz M, Palme K, Simons M, Diester I, Brox T, Ronneberger O: U-net: deep learning for cell counting, detection, and morphometry. *Nat Methods* 16:67–70, 2019
30. Hodneland E, Kögel T, Frei DM, Gerdes HH, Lundervold A: CellSegm - a MATLAB toolbox for high-throughput 3D cell segmentation. *Source Code Biol Med* 8:16, 2013
31. Gautama S, Goeman W, D’Haeyer J: Robust detection of road junctions in vhr images using an improved ridge detector. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol 35, 2004, pp 815–819
32. Cortes C, Vapnik V: Support-vector networks. *Machine learning* 20(3):273–297, 1995
33. Gonzalez RC, Woods RE, Eddins SL: *Digital Image Processing Using MATLAB* London: Pearson Education, 2004
34. Ubul K, Yadikar N, Amat A, Aysa A, Yibulayin T: Uyghur document image retrieval based on gray gradient co-occurrence matrix. In: *2015 Chinese Automation Congress (CAC)*, 2015, pp 762–766
35. Dietterich TG: Ensemble methods in machine learning. In: *Multiple Classifier Systems*. Springer, Berlin, 2000, pp 1–15
36. Walpole RE, Myers RH: *Probability and Statistics for Engineers and Scientists* Englewood Cliffs: Prentice-Hall, 1993

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.