# Corneal Endothelial Cell Segmentation by Classifier-Driven Merging of Oversegmented Images

Juan P. Vigueras-Guillén, Eleni-Rosalina Andrinopoulou, Angela Engel, Hans G. Lemij, Jeroen van Rooij, Koenraad A. Vermeer, *Member, IEEE*, and Lucas J. van Vliet, *Member, IEEE*

***Abstract*— Corneal endothelium images obtained by *in vivo* specular microscopy provide important information to assess the health status of the cornea. Estimation of clinical parameters, such as cell density, polymegethism, and pleomorphism, requires accurate cell segmentation. State-of-the-art techniques to automatically segment the endothelium are error-prone when applied to images with low contrast and/or large variation in cell size. Here, we propose an automatic method to segment the endothelium. Starting with an oversegmented image comprised of superpixels obtained from a stochastic watershed segmentation, the proposed method uses intensity and shape information of the superpixels to identify and merge those that constitute a cell, using support vector machines. We evaluated the automatic segmentation on a data set of *in vivo* specular microscopy images (Topcon SP-1P), obtaining 95.8% correctly merged cells and 2.0% undersegmented cells. We also evaluated the parameter estimation against the results of the vendor's built-in software, obtaining a statistically significant better precision in all parameters and a similar or better accuracy. The parameter estimation was also evaluated on three other data sets from different imaging modalities (confocal microscopy, phase-contrast microscopy, and fluorescence confocal microscopy) and tissue types (*ex vivo* corneal endothelium and retinal pigment epithelium). In comparison with the estimates of the data sets' authors, we achieved statistically significant better accuracy and precision in all parameters except pleomorphism, where a similar accuracy and precision were obtained.**

***Index Terms*— Specular microscopy, confocal microscopy, merging superpixels, stochastic watershed, support vector machines.**

## I. Introduction

THE corneal endothelium (CE) is a single layer of closely packed and predominantly hexagonally-shaped cells forming the inner surface of the cornea. It performs an important role maintaining an optimal state of hydration of the cornea [1]. Endothelial cell loss occurs normally due to aging, from a density of approximately 5,000 cells/mm$^2$ at birth down to 2,000 cells/mm$^2$ in a normal adult eye [2]. However, this loss can be exacerbated by trauma, diseases, or intraocular surgery, which might in turn lead to irreversible corneal swelling when the cell density drops below 500 cells/mm$^2$ [3], [4]. Due to its limited cell division capacity, the repair function is usually restricted to the swelling and sliding of the existing cells in order to maintain the barrier properties. Quantitative analysis of the corneal endothelial cell morphology from image data provides clinical information to assist ophthalmologists in diagnosis and treatment of corneal diseases. Currently, endothelial cell density (ECD) is the most important measure to assess the corneal health state, reported as the number of cells per mm$^2$. Other parameters such as polymegethism – expressed by the coefficient of variation (CV) in cell size – or pleomorphism – quantified by the hexagonality coefficient (HEX) as the percentage of hexagonal cells – are not used in standard clinical practice due to unreliable estimations.

Different instruments are available to assess the cell density and morphometry of the endothelium. Non-contact specular microscopy, developed in the late 1970's, is a fast, non-invasive, *in vivo* imaging method, which permits to record large field-of-view images of the endothelium. In contrast, contact confocal microscopy, developed in the late 1980's, provides an *in vivo*, high-resolution assessment of all corneal

Manuscript received February 7, 2018; revised May 25, 2018; accepted May 25, 2018. Date of publication May 29, 2018; date of current version October 1, 2018. This work was supported by the Dutch Organization for Health Research and Health Care Innovation (ZonMw) under Grants 842005004 and 842005007. *(Corresponding author: Juan P. Vigueras-Guillén.)*

J. P. Vigueras-Guillén is with the Rotterdam Ophthalmic Institute, 3011 BH Rotterdam, The Netherlands, and also with the Department of Imaging Physics, Delft University of Technology, 2611 LC Delft, The Netherlands (e-mail: j.p.viguerasguillen@tudelft.nl).

E.-R. Andrinopoulou is with the Rotterdam Ophthalmic Institute, 3011 BH Rotterdam, The Netherlands, and also with the Department of Biostatistics, Erasmus MC, 3015 CN Rotterdam, The Netherlands (e-mail: e.andrinopoulou@oogziekenhuis.nl).

A. Engel, and K. A. Vermeer are with the Rotterdam Ophthalmic Institute, 3011 BH Rotterdam, The Netherlands (e-mail: a.engel@oogziekenhuis.nl; k.vermeer@oogziekenhuis.nl).

H. G. Lemij and J. van Rooij are with the Rotterdam Eye Hospital, 3011 BH Rotterdam, The Netherlands (e-mail: h.lemij@oogziekenhuis.nl; j.vanrooij@oogziekenhuis.nl).

L. J. van Vliet is with the Department of Imaging Physics, Delft University of Technology, 2611 LC Delft, The Netherlands (e-mail: l.j.vanvliet@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2018.2841910

layers, but requires physical contact with the cornea and relatively long acquisition time [5]. Due to its non-invasive nature and fast acquisition time, specular microscopy is the current clinical standard technique. While contact confocal microscopy obtains considerably clearer images from edematous corneas [6], both microscopes provide images of sufficient quality from the central and peripheral cornea. Due to the optical principles of the technology, good quality specular microscopy images are limited to corneas that have a smooth endothelium surface [7], [8]. Non-contact confocal microscopy is the most recent modality (early 2000s), and it provides similar image quality as the non-contact specular microscopy for normal corneas but yields a larger field-of-view [9].

Manual quantification of endothelial images is very labor-intensive, so computer aided techniques to generate the cell segmentation are necessary. Microscope manufacturers currently provide built-in software to automatically segment the recorded images and estimate the parameters. Several studies using different microscope modalities have been performed in order to evaluate the reliability of these fully automated analyses in comparison with semi-automated and manual analyses, especially in the evaluation of cell density. In general, the fully automatic ECD estimations were not consistent with the manual estimations. Some studies indicated an overestimation of ECD in both specular and confocal microscopy images for healthy and glaucomatous eyes exhibiting a high cell density [9], and for healthy and transplanted corneas with normal to low cell density [10], [11], mainly due to the oversegmentation of large cells. In contrast, another study (employing a different microscope vendor's software) reported an underestimation of ECD in both, confocal and specular microscopy, for normal corneas with high cell density due to the identification of false cell borders [12]. When images with large differences in cell density were analyzed with the same software, this was prone to underestimating the cell density in images with high ECD and overestimating it in images with low ECD [13], or vice versa [7]. Previous studies on semi-automated analyses reported clinically acceptable results as it produces ECD values comparable to the manual analysis. However, this process is still time consuming since an expert needs to correct the segmentation.

Overall, the automatic detection of cell borders is a complicated task due to the presence of noise, variation in illumination, and optical artifacts. Furthermore, the estimation of polymegethism and pleomorphism can be significantly affected by just a few errors in the segmentation. Previous studies suggest that the built-in software of the currently available commercial microscopes is prone to mistakes in the segmentation, particularly in the presence of a low or high cell density or a high degree of polymegethism. Therefore, there is a need to develop an algorithm that can reliably estimate these parameters from images with such characteristics.

### A. Related Work

Several cell segmentation techniques for corneal endothelium images have been proposed. Selig *et al.* [14] suggested a stochastic watershed approach to segment *in vivo* confocal microscopy images. Ruggeri *et al.* segmented *ex vivo* porcine endothelium images obtained with inverse contrast phase microscopy by means of an artificial neural network algorithm and *in vivo* confocal microscopy images by means of a genetic algorithm [16]. Gavet and Pinoli [17] proposed a binarization algorithm to segment *in vivo* specular microscopy images. Finally, Sharif *et al.* [18] developed a hybrid model based on snake and particle swarm optimization for *in vivo* confocal microscopy images.

### B. Our Contribution

Due to the aforementioned challenges, i.e., illumination distortions, optical artifacts, and presence of noise, current segmentation methods fail to identify all cell edges without finding invalid ones. These methods seem to be designed to segment a specific type of endothelium images within a limited range of cell density and variation of cell size. Here, we propose a machine learning framework based on Support Vector Machines (SVM) for segmenting a broad range of endothelium images from different types of microscopes. We hypothesize that, by starting with an oversegmented image where all cell edges have been traced, a trained classifier could identify the fragments (superpixels) that constitute a complete cell. Indeed, the union of fragments belonging to the same cell shows distinctive features (related to shape, intensity, size, etc.) compared to an arbitrary combination of fragments from different cells. Thus, by merging those fragments, the oversegmented image converges towards the correct segmentation. The creation of the initially oversegmented image is not a trivial task, and here we adapted Selig *et al.*'s approach [14] to generate an initial segmentation without undersegmentation and where hardly any cell is divided in more than three fragments.

Although superpixel merging methods have been proposed in recent years, they were mainly applied to high quality color images and using color histogram features [19], [20], a combination of color and spectral information [21], or color with superpixel size information [22]. Needless to say, all these methods fail to provide acceptable results in specular microscopy images due to the challenges posed by the low image quality. In contrast, the novelty of our method lies in exploiting the morphology of the endothelial cell layer in addition to intensity information of the constituting superpixels and the candidate merger, as well as the edges between them. Moreover, we have exploited the idea that better segmentation results can be achieved if, instead of aiming directly for an optimal segmentation, an oversegmentation followed by a merging process is employed.

To evaluate our method, we used a dataset of endothelial images obtained with *in vivo* specular microscopy (Topcon SP-1P) and we compared our results against the estimates of the microscope's built-in software. To illustrate the versatility of the method, it was also applied to three other datasets: 52 confocal microscopy endothelial images used in Selig *et al.*'s paper [14], 30 *ex vivo* endothelial images from phase-contrast microscopy published in Ruggeri *et al.*'s paper [15], and a dataset of 23 confocal microscopy fluorescence images of the central retinal pigment epithelium (RPE)

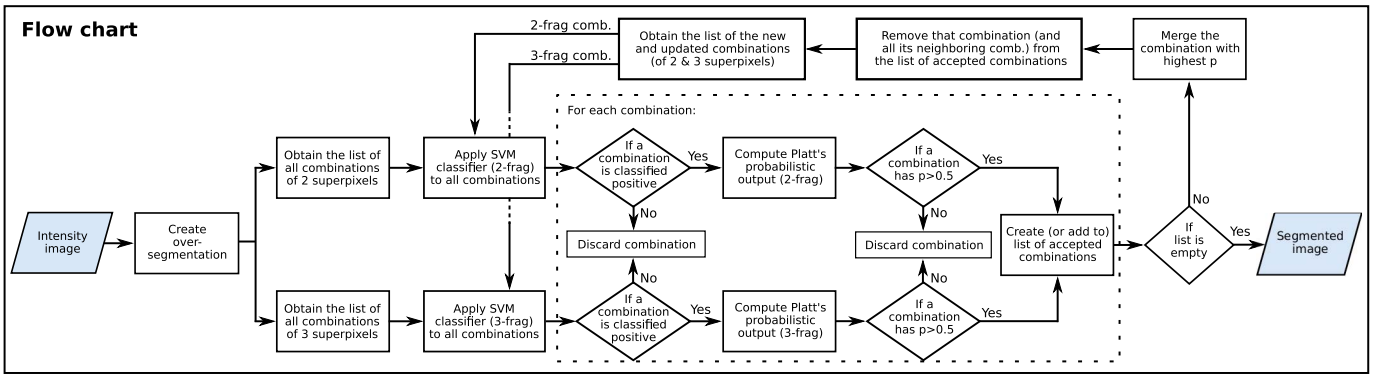



Fig. 1. Flow chart of the method.

analyzed in Chiu *et al.*'s paper [23]. RPE cells might show a considerably high variation in cell size, unlikely in corneal endothelial cells, which makes it a good example to test the robustness of the method against polymegethism.

This paper is organized as follows. Section II provides a description of the proposed method, including how the initial oversegmentation was generated. Section III first evaluates the merging process and the subsequent cell segmentation of the main dataset. Second, it evaluates the estimates of the clinical parameters for all datasets. Third, a brief explanation about computational cost is provided. Finally, the results are discussed in Section IV.

## II. Methodology

### A. Method Summary

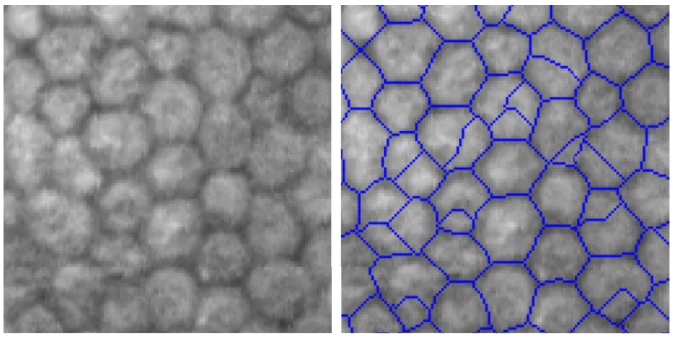
An SVM algorithm is presented, which aims to create a correct segmentation of an endothelial cell image from an initially oversegmented image composed of superpixels. This is accomplished by evaluating whether combinations of adjacent superpixels constitute a complete cell. The SVM classifier uses information about shape, structure, and intensity from the superpixels (both, separately and combined) and their boundaries. All possible combinations of two and three adjacent superpixels in the image are considered simultaneously by means of a dedicated classifier for each type of combination (2 or 3 fragments). In order to compare the combinations from the different classifiers, the signed distance from the classification hyperplane is transformed into a probabilistic output by means of Platt's algorithm [24]. Only the combinations classified as positives and with probability higher than 0.5 are considered. The combination with the highest probability is merged, iterating this process until no more combinations are accepted.

To optimize this iterative process, we create a list of 'accepted combinations' such that every time a merger is performed and removed from the list, we also remove all combinations that are no longer possible and all combinations that need to be updated (i.e., neighboring combinations to the merger whose features have changed). Then, all new combinations involving the new superpixel are evaluated as described above, as well as the neighboring combinations that needed to be updated. These combinations are added to the list if their mergers are accepted. Hence, only the strictly necessary (or new) combinations are re-evaluated after each merger, making the process computationally efficient. Once the list of accepted combinations is empty, we have reached the final segmentation. A data-flow diagram is depicted in Fig. 1.

We chose an SVM classifier with a probabilistic output instead of any other type of classifier or a different method (clustering, regression, etc.) because of the following reasons: (1) SVMs belong to the class of supervised techniques, which usually provide better results than unsupervised methods such as clustering techniques; (2) SVMs can use the features of the constituting superpixels (independently) in combination with the features of the candidate merger in order to determine if a merger is correct, which makes the merging decision more robust; (3) SVM's regularization parameter ($C$) can be used to avoid over-fitting; (4) SVMs are well known to provide great performance in tasks with highly-unbalanced classes, which happens in our problem; (5) SVMs are robust against the high-dimensional space of the problem; (6) SVM's kernel trick allows building a more complex decision boundary, and in our case it also helps to deal with the arbitrary order of the superpixels in the feature vector; (7) SVMs are computationally very efficient (fast in evaluating), which is necessary in clinical applications with thousands of combinations to be evaluated in one single image; (8) Platt's probabilistic output permits comparing competing types of combinations (2- and 3-fragments), which cannot be done directly.

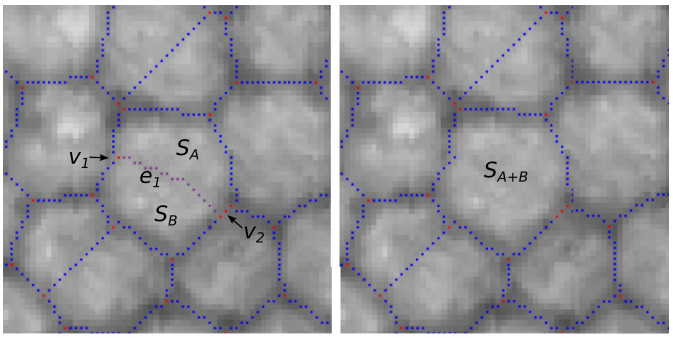
### B. Initial Oversegmentation

In an ideal oversegmented image, all cell edges are traced with high accuracy and each cell is comprised by a minimum number of superpixels. Due to the low signal-to-noise ratio of specular microscopy images, these two goals are difficult to achieve simultaneously. Since only the removal of edge segments is considered by our method, any undetected cell edge (undersegmentation) produces an error that cannot be corrected *a posteriori*. On the other hand, creating a highly oversegmented image might trace all cell edges accurately, but makes the subsequent model too complex and prone to mistakes.



(a) (b)

Fig. 2. (a) Portion of a specular microscopy image. (b) Initial oversegmentation (in blue). Cells are comprised of 1, 2, or 3 superpixels. In the specular microscopy dataset, rarely any cell appears divided in more than 3 fragments.




(a) (b)

Fig. 3. Visual representation of a 2-fragment merger. The segmentation (edges in blue, vertices in red) is superimposed on the intensity image. (a) Indication of the elements involved in the merger. (b) After the merger, vertex $v_1$ is transformed into an edge-pixel, while $v_2$ stays as a vertex.

Selig *et al.* [14] proposed a method to segment endothelial images by applying a seeded watershed algorithm in a stochastic manner. The seed grid has a hexagonal pattern with a spacing equal to the most common cell diameter in the image ($l$), which is computed as the inverse of the characteristic frequency, $l = 1/f^*$, estimated by Fourier analysis [14]. By creating a denser grid of seeds, we can control the degree of oversegmentation. Here, we divided the estimated cell diameter by $\sqrt{3}$, which corresponds to a grid with three times more seeds, $n_{seeds} = A_I(f^*\sqrt{3})^2 = 3A_I f^{*2}$, being $A_I$ the total image area [14]. This enabled a good detection of all cells edges without creating an excess of oversegmentation (Fig. 2).

For the remaining parameters in the stochastic watershed method (iterations, added noise, blur size, and local minima to ignore), we used the values reported in the original paper [14]. One of these parameters, $k_\sigma = \sigma_{PDF} f^* = 0.17$, is related to a Gaussian smoothing filter ($\sigma_{PDF}$) that is applied to the stochastic watershed output. Selig *et al.* observed that, for large $\sigma_{PDF}$, the ridges in the segmentation were shifted with respect to those in the input image. To solve this, Selig *et al.* added a final, optional step where the classic seeded watershed method was applied to a smoothed version of the input image and the segmented regions of the stochastic watershed were used as seeds. Only two datasets evaluated in this paper showed such a shift in the ridges, namely Selig *et al.*'s confocal microscope data ($\sigma_{PDF} \approx 6$) and the fluorescence confocal RPE data ($\sigma_{PDF} \approx 4.5$). For the other two datasets ($\sigma_{PDF} \approx 3$), this final step did not substantially change the segmentation (phase-contrast microscope data) or it even degraded the segmentation in areas with blurry cell edges (specular microscope data). Therefore, the edge correction method was only applied to the confocal microscopy datasets.

### C. Definition of a Merger

Given a segmented (binary) image, we define the vertices as the branch points of the segmentation, and the edges as the set of 8-connected positive pixels whose endpoints are constrained to vertices. Biologically, a vertex is placed in the point where the cell edges from three or more cells meet. In order to avoid small edge segments comprised by less than two pixels, these are fused with their respective vertices to form a single vertex (Fig. 3a, vertex $v_2$). The superpixels are defined as the sets of 4-connected negative (non-edge) pixels. We define a 2-fragment merger as the union of two adjacent superpixels by the removal of the edge segment between them. A 3-fragment merger is defined as the union of three adjacent superpixels by the removal of the two or three edge segments between them. This is accomplished by two consecutive 2-fragment mergers.

### D. Features

For every candidate merger, let us consider $S_A$ and $S_B$ as the two fragments to be merged, $S_{A+B}$ as the resulting superpixel, $e_1$ as the edge segment that needs to be removed to create $S_{A+B}$, and $v_1$ and $v_2$ as the vertices at the ends of $e_1$ (Fig. 3a). A merger is computed as $S_{A+B} = S_A + S_B + e_1$, and $v_1$ with $v_2$ are evaluated to identify whether they stay as vertices or they become part of the edge of $S_{A+B}$ (Fig. 3b). The following features are computed for each superpixel in the merger ($S_A$, $S_B$, and $S_{A+B}$). Let $S$ be any of those superpixels:

1) Normalized cell size, computed as the number of pixels in $S$ (area) divided by the size, in pixels, of the most common cell (estimated by Fourier Analysis [14]).
2) Normalized cell intensity, computed as the ratio of the average pixel intensity of $S$ to the average pixel intensity of the neighboring superpixels.
3) Convexity, defined as the ratio of the perimeter of the convex hull of $S$ to the perimeter of $S$.
4) Solidity, defined as the proportion of the pixels in the convex hull that are also in $S$ ($Area/ConvexArea$).
5) Circularity, defined as $4\pi\, Area/Perimeter^2$.
6) Eccentricity, defined as the eccentricity of the ellipse that has the same second-order moments as $S$.
7) Number of neighboring superpixels.
8) Standard deviation of the lengths of the edges associated with $S$.
9) Standard deviation of the inner angles. We define the inner angles as the angles that are formed in the center

of mass of $S$ between the straight lines that are traced from that point to the vertices of $S$.

10) Standard deviation of the outer angles. We define the outer angles as the angles formed in the vertices of $S$ between the straight lines that are traced between the vertices.
11) Ratio of the area of $S$ to the polygon area, where the polygon is formed by connecting the vertices of $S$ with straight lines.
12) Binary value to indicate whether $S$ lies in the border region of the image (this is not included for $S_{A+B}$).

For $S_{A+B}$, we compute extra features. Here, we define $l_1$ as the straight line that is formed by connecting $v_1$ with $v_2$. Features related to those vertices provide two values, one per vertex.

13) Ratio of the average intensity of $e_1$ to the average intensity of $S_A$ and $S_B$.
14) Ratio of the average intensity of $e_1$ to the average intensity of the edges of $S_{A+B}$.
15) Ratio of the area formed between $e_1$ and $l_1$ to the area of $S_{A+B}$.
16) Ratio of the length of $l_1$ to the length of $e_1$.
17) For each vertex ($v_1$, $v_2$), ratio of the average intensity of half $e_1$ (the part closest to the vertex) to the average intensity of the edges of $S_{A+B}$ ending in $v$ (only the half of the edges closest to $v$). In contrast to feature 14, here only the intensity in the vicinity of $v$ is evaluated (2 values).
18) Ratio of the length of $e_1$ to the average edge length of $S_A$ and $S_B$ (2 values).
19) Standard deviation of the angles formed in the vertices ($v_1$ and $v_2$) before the merger occurs (2 values).
20) The angles formed in the vertices ($v_1$ and $v_2$) after the merger occurs (2 values).

The concatenation of the features of the involved superpixels creates the feature vector of a merger. For an $n$-fragment merger, $n!$ possible feature vectors can describe the same merger, depending on the order of the fragments. The SVM kernel was designed to map all those vectors to the same point in feature space (see Section II-F).

For a 2-fragment merger, the feature vector contains 47 elements. For a 3-fragment merger, the extra features are computed twice (one per removed edge segment), and they are processed as follows: For features 13-16, the averaged value is computed; for features 17-20, the maximum and minimum values are kept. In total, a 3-fragment merger is represented with a feature vector of 59 elements.

### E. Class Definition & Training Data Preparation

In the 2-fragment SVM classifier, the *good-merger* class is defined as the set of combinations of two superpixels that form a cell. The *bad-merger* class is defined as the set of all other combinations of two adjacent superpixels, with the exception of combinations coming from within cells divided in more than two fragments. The classes in the 3-fragment SVM classifier are defined in an analogous way, but now for sets of three superpixels.

To prepare the training/test data, we retrieved the class elements from the oversegmented images as described above. Since some features are affected by the state of fragmentation of the neighboring cells, the training elements in the good-merger class were computed twice, once when none of the fragmented neighboring cells were merged yet, and again when all were merged. Bad-merger examples were also retrieved from the resulting superpixel of a good-merger with its neighboring superpixels. Hence, the training/test examples not only included all the mergers that could be formed in the initially oversegmented image, but also (some of) the mergers that could appear during the merging process.

To help solve the segmentation of cells divided in more than three fragments without having a classifier specifically built for that purpose, the following training elements in the good-merger class were included: for the $n$-fragment classifier, a cell divided in $m$ fragments, being $m > n$, will generate combinations of $n$ superpixels where one superpixel is the result of previously merging $m - n + 1$ fragments of the cell.

To deal with cells touching the border of the image, some good and bad mergers were included for those cases. Feature no 12 was added for this purpose. Eventually, the estimation of the endothelial parameters is done excluding the superpixels touching the border region of the image.

The ratio of bad-mergers (negative class) to good-mergers (positive class) in the training/test dataset was around 25:1 for the 2-fragment classifier, and 200:1 for the 3-fragment classifier.

### F. SVM, Kernel, & Probabilistic Model

By design, no oversampling or undersampling is applied to the classes, and each element in any of the two classes has the same weight. For computational purposes, the negative class in the 3-fragment classifier was undersampled to the same ratio as in the 2-fragment classifier without losing discriminative power. This was done by selecting the closest negative elements to the positive ones, in Euclidean distance, which removes elements that exert no influence on the SVM hyperplane.

A Gaussian-based kernel was designed, which maps all permutations of the feature vectors of a single merger onto the same point in the feature space. Let us suppose $\mathbf{G}(\mathbf{x}_j, \mathbf{x}_k)$ denotes element $(j, k)$ in the Gram matrix, where $\mathbf{x}_j$ and $\mathbf{x}_k$ are $p$-dimensional vectors representing elements $j$ and $k$. For the 2-fragment SVM classifier, let $j$ indicate the merger between superpixels $A$ and $B$, whose vector can be depicted as $\mathbf{x}_j = [f_A, f_B, f_{A+B}]$ or $\mathbf{x}'_j = [f_B, f_A, f_{A+B}]$, with $f_n$ the features of the superpixel $n$. Then, the kernel is defined as

$$\mathbf{G}(\mathbf{x}_j, \mathbf{x}_k) = \exp(-\gamma\,||\mathbf{x}_j - \mathbf{x}_k||^2) + \exp(-\gamma\,||\mathbf{x}'_j - \mathbf{x}_k||^2), \quad (1)$$

with $\gamma$ the scale parameter of the Gaussian radial basis function. The kernel for the 3-fragment classifier was defined in a similar fashion, covering the six possible permutations that describe the same 3-fragment merger.

In order to compare mergers from different SVM classifiers, the optimal score-to-posterior-probability transformation was computed for each classifier by using Platt's approach [24].

## III. Experiments & Evaluation

We evaluate the proposed method in Section III-A for the main dataset, which is described in Section III-A1. The evaluation is done for the classifiers separately in Section III-A2, for the classifiers jointly in Section III-A3, and for the full cell segmentation task in Section III-A4. In Section III-B we evaluate the estimates of the clinical parameters from the segmented images for all datasets, which are described in Section III-B1. The estimates are compared against the gold standard and either the estimates of the microscope's built-in software or the estimates provided by the original authors.

For all datasets, a double 5-fold cross-validation approach was used to separate training, validation, and testing sets, providing a ratio in the size of the sets of approximately 64% training, 14% validation, and 20% testing. Images were completely assigned to a single fold to prevent combinations of fragments from the same image in different sets. For the evaluations in Section III-A, the experiments were repeated five times, reassigning the images to the folds differently each time, and the averaged results were reported.

### A. Evaluation of the Method

*1) Dataset:* The main dataset consists of 30 corneal endothelium images from the central cornea of 30 glaucomatous eyes, with approximately 250 cells per image. They were obtained with a non-contact specular microscope (SP-1P, Topcon Co, Japan) for an ongoing study regarding the implantation of a Baerveldt glaucoma drainage device in The Rotterdam Eye Hospital. The images were acquired prior to the implantation of the device. They cover an area of 0.25 mm $\times$ 0.55 mm and were saved as 241$\times$529 pixels 8-bits grayscale images. The acquisition occurred with informed consent and followed the tenets of the Declaration of Helsinki. One expert created the gold standard for each image by performing manual segmentation of the cell contours using an open-source image manipulation program (GIMP v.2.8).

*2) Evaluating the Accuracy of the Classifiers:* The performances of the two classifiers (2- and 3-fragments) were evaluated independently. Fifteen SVMs were computed per classifier by varying the misclassification cost between False Positives (FP) and False Negatives (FN). The cost values FP:FN followed the ratio $5^{n/2}$:1, with $n = 12, 11, \ldots, 1, 0, -1, -2$. Based on the class definition, a FP is defined as a bad-merger classified as positive, whereas a FN is a good-merger classified as negative. A grid search was performed to find the best SVM parameters, $\gamma$ and $C$, in the validation set, using the cost function: $\text{argmin}_{\gamma,C}\,(cost_{FP}\text{FP} + cost_{FN}\text{FN})$.

The evaluation was done on the test set, built as described in Section II-E. FROC curves [25] were generated by calculating the true positive rate, TPR = TP/(TP+FN), and the fractional number of false positives per cell, FPs/cell, from the output of each classifier. Two setups were considered: one where the fifteen classifiers performed the task independently, and another for a cascade of classifiers. In the latter, the classifiers were ordered from highest to lowest ratio of cost FP:FN, and any accepted merger at one classifier was automatically accepted by the following classifiers. The classifier with equal costs (FP:FN of 1:1) was marked in all curves (Fig. 4a-b).

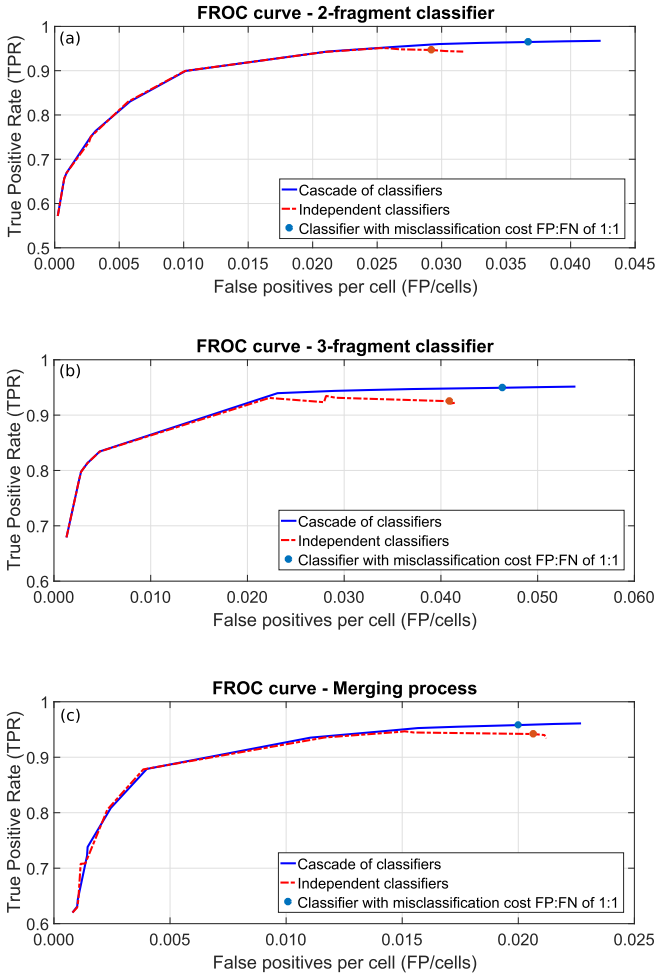



Fig. 4. FROC curves of the 2-fragment classifier (a), 3-fragment classifier (b), and the outcome of the merging process (c), using 15 classifiers with different costs of FP:FN. The 15 classifiers are evaluated in two ways: independently and connected in a cascade. In all cases, the classifier with equal cost (FP:FN of 1:1) is marked.

An inflection point was observed in the curves of the independent classifiers, at those with costs 5:1 (2-frag) and 25:1 (3-frag). Beyond that point, decreasing the misclassification cost of the FP created more FPs, but did not increase the number of TPs. In contrast, the cascade of classifiers benefited from a higher TPR because of the accumulative effect of TPs in a cascade design, but at the expense of a slightly higher number of FPs. It was clear from the graphs that no further improvement occurs beyond the classifier with equal costs (FP:FN of 1:1), as the remaining good-mergers would come at a high price in terms of FPs. At that point (1:1), the 2-fragment classifier in the cascade provided a TPR of 96.5% with 0.037 FPs/cell, and the 3-fragment classifier provided a TPR of 94.9% with 0.046 FPs/cell. Moreover, half of the FPs in both cases came from cells touching the image borders and such cells are not considered when calculating ECD, CV, or HEX.

*3) Evaluating the Accuracy of the Merging Process:* This evaluation differed from the previous one in three aspects: (1) the setup follows the proposed method (Fig. 1), which means that both classifiers work simultaneously; (2) the evaluation

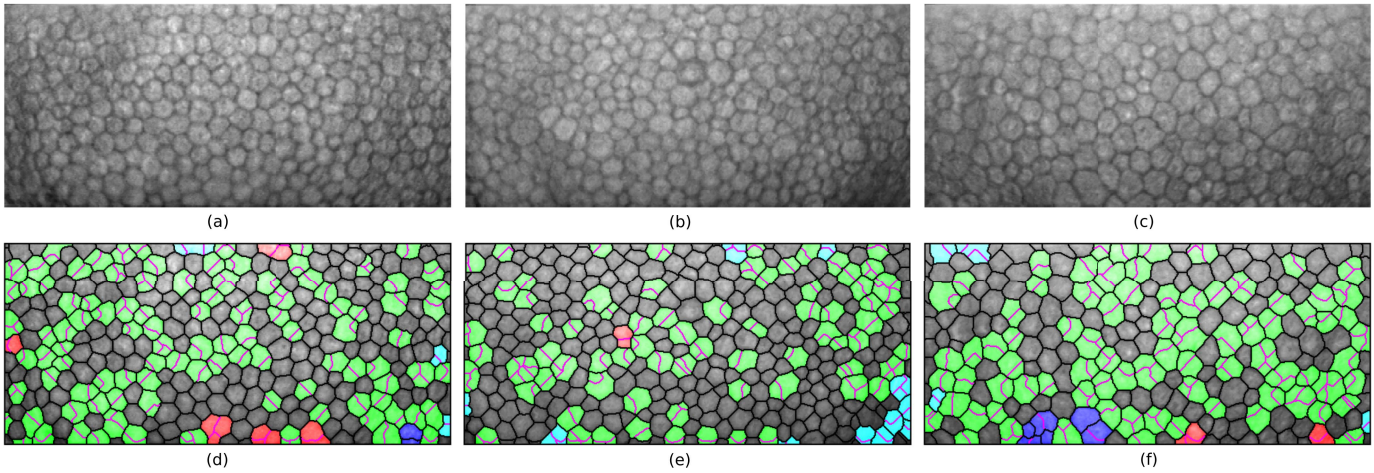



Fig. 5. Three representative specular microscopy images (a)-(c), and their respective visual outcome of the merging (d)-(f): black lines indicate the segmentation outcome, and magenta lines indicate edges being removed in the merging process. All fragments involved in mergers are colored: green denotes complete cells after merging, red denotes undersegmented cells (FP), dark blue denotes oversegmented cell (FN), and light blue denotes inconclusive mergers.

now uses the test images, which means that only mergers that occur during the merging process are involved; (3) the evaluation metric is now defined for the final result at the cell level: a TP is defined as a fragmented cell that is correctly merged, a FN is defined as a fragmented cell that is not completely merged (oversegmentation), and a FP is defined as any cell âŁ"- fragmented or not âŁ"- that suffers from an incorrect merger (undersegmentation). A cell can be labeled with only one condition, and the condition FN prevails over FP. For instance, if a fragment from an oversegmented cell (FN) merges with another cell, the former cell is labeled as FN and the latter as FP. If a fragmented cell, after merging correctly, merges with another cell, then both are labeled as FP. Hence, TPR indicates the percentage of oversegmented cells that were correctly merged, whereas the number of FPs/cell indicates the percentage of cells that were undersegmented. This metric will be also used for the remaining datasets.

The same fifteen SVMs were used here, and both setups, independent classifiers and the cascade of classifiers, were considered. In the latter, once all the accepted combinations have been merged in one set of classifiers, the output segmentation is provided as input to the next set of classifiers. The FROC curves (Fig. 4c) showed a similar pattern as before. The cascade, whose main goal is to merge the most certain true combinations in the first stages, could potentially reduce FPs in the last stages as less false combinations are being evaluated, showing a small but clear effect. At the classifier with equal costs (FP:FN of 1:1) in the cascade, a TPR of 95.8% with 0.020 FPs/cell was obtained. To illustrate the outcome, at that point there are, on average, less than six undersegmented cells and less than two oversegmented cells per image (Fig. 5), considering that initially 25% of the approximately 250 cells per image were oversegmented. Furthermore, the majority of undersegmented cells were located along the image borders. This supports the idea of having reached the optimal point, since the number of over- and undersegmented cells in the inner part of the image was similar, while undersegmentation was starting to prevail along the image borders.

*4) Evaluating the Accuracy of the Segmentation:* The segmentation accuracy was measured by using the distance transform. In a first measure, the distance transform was computed on the resulting segmentation. The distance values were averaged over those pixels that were labeled as edge pixels in the gold standard, providing a single value per image. This measure is sensitive to undersegmentation, as missing true edges in the output segmentation increases the value, but is barely affected by oversegmentation. In a second measure, the distance transform was computed on the gold standard, retrieving only the values of those pixels that were labeled as edge pixel in the resulting segmentation. In contrast to the former measure, this is sensitive to oversegmentation, as false edges increase the value, but insensitive to undersegmentation. In almost all images, the first measure was slightly higher (the difference was $0.05 \pm 0.05$, mean $\pm$ SD), suggesting a very small dominance of undersegmentation. The average of both measures was computed for each image, providing a value of $0.93 \pm 0.09$ (mean $\pm$ SD). This suggested that the segmentation was on average within one pixel from the gold standard.

## B. Evaluation of the Application

Based on the previous results, a cascade of four classifiers was established as the final setup, with cost values (FP:FN) of 125:1, 25:1, 5:1 and 1:1.

*1) Datasets & Statistical Analysis Method:* Besides the main dataset (Section III-A1), three other, publicly available datasets were evaluated.

The confocal microscopy dataset includes 52 *in vivo* images (768×576 pixels, Nidek Confoscan 4, Italy) of transplanted corneal endothelium (DSAEK), imaged one year after surgery. Images were cropped by the author to remove dark areas. Each image contains approximately 70 cells [14].

The phase-contrast microscopy dataset includes 30 *ex vivo* images (768×576 pixels, Olympus CK 40, USA) of porcine corneal endothelium. These images were also cropped by the author. Each image contains approximately 350 cells [15].

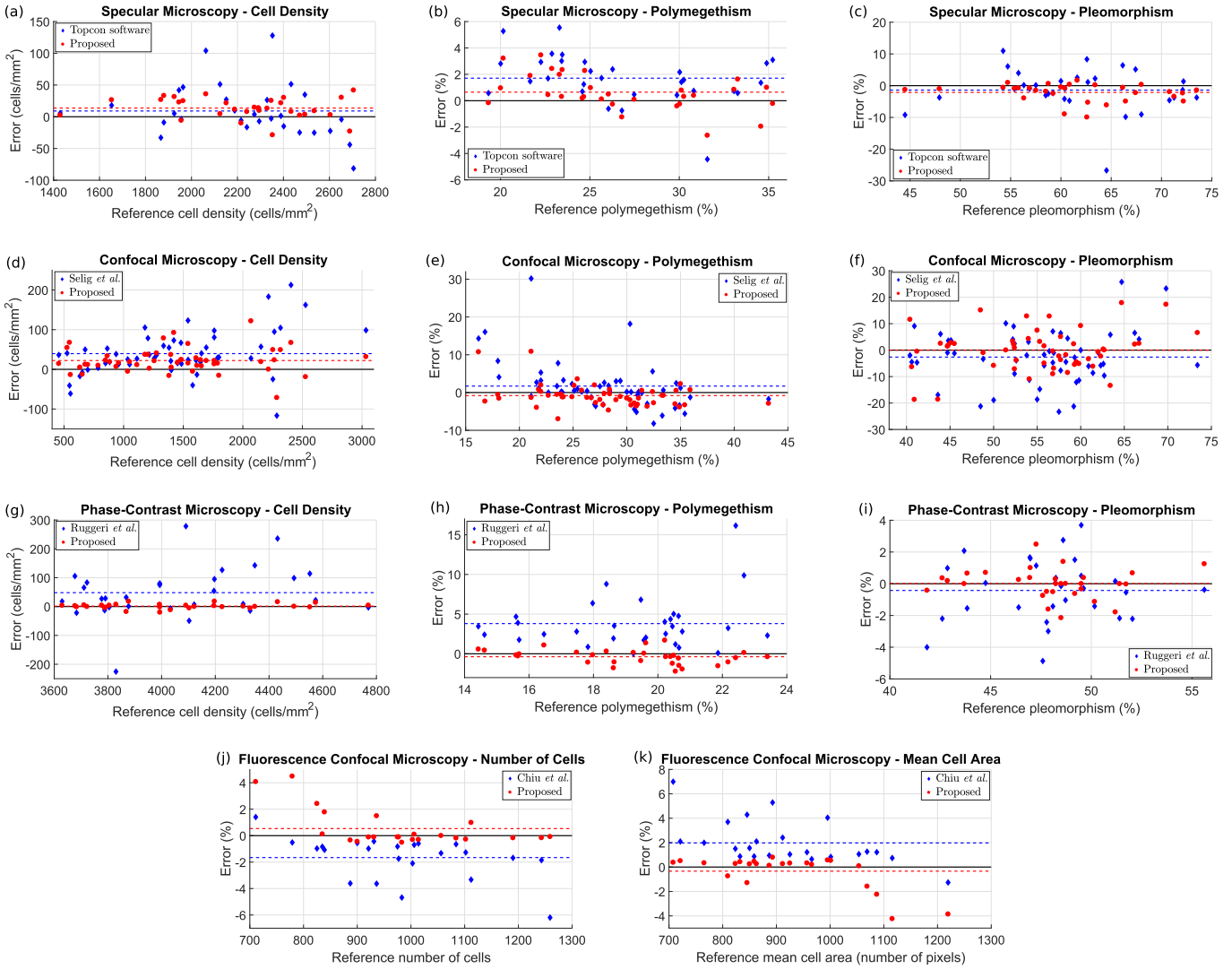



Fig. 6. Estimates of the clinical parameters in all datasets: Specular microscopy (a)-(c), confocal microscopy (d)-(f), phase-contrast microscopy (g)-(i), and fluorescence confocal microscopy (j)-(k). The x-axis indicates the value for the gold standard, and the y-axis indicates the error computed either as the difference between the estimates and the gold standard estimates (a)-(i), or the percentage of such error (j)-(k). Each point corresponds to one image in the dataset (proposed results in red, third-party results in blue). The mean value of the error for each set is drawn with a discontinuous line.

The fluorescence confocal microscopy dataset includes 23 images (1024×1024 pixels, Nikon Eclipse C1, Japan) of central RPE from 17 mice, with approximately 1000 cells per image [23].

The parameters to be estimated were endothelial cell density (ECD), polymegethism (CV), and pleomorphism (HEX) for the endothelium datasets, and the number of cells and the average cell size for the RPE images (as it was done in the original papers). The gold standard and the segmentation images produced by the original authors were publicly available or provided at our request, which allowed us to apply the same algorithm for parameter estimation in all cases. For all images in all datasets, only the cells covered by the area of the gold standard were included for the parameter estimation. The only exception was Topcon's segmentation (specular microscopy dataset), since the microscope's software did not provide any cell segmentation in areas where cells were not correctly detected by the software. In that dataset, the gold standard covered a surface 35% larger. The estimation error was defined as the difference between the estimated value and the gold standard value for the CE datasets (Fig. 6a-i), and as the percentage of that difference for the RPE dataset (Fig. 6j-k). Note that, for polymegethism (Fig. 6b, 6e, 6h) and pleomorphism (Fig. 6c, 6f, 6i), the parameter values are provided as a percentage, where the error is the difference of the percentages. The mean value and SD of those estimation errors are indicated in Table I.

The RPE images and their gold standard were originally presented by Ding *et al.* [26], but Chiu *et al.* [23] discovered (but did not correct) in the gold standard a few unsegmented cells due to their small size; here those cells were corrected (on average, three cells per image were added). We also noticed that cell edges along the perimeter of the gold standard area were not annotated, and they were corrected. For the

TABLE I
MEAN AND STANDARD DEVIATION OF THE ESTIMATION ERROR OF THE CLINICAL PARAMETERS FOR ALL DATASETS

| Dataset | ECD (cells/mm$^2$) | CV (%) | HEX (%) |
|---|---|---|---|
| Specular Microscopy | | | |
| Proposed | $14 \pm 16$ | $0.6 \pm 1.3$ | $-2.1 \pm 2.7$ |
| Topcon | $9 \pm 41$ | $1.7 \pm 1.8$ | $-1.4 \pm 6.8$ |
| Confocal Microscopy | | | |
| Proposed | $22 \pm 30$ | $-0.8 \pm 3.0$ | $0.0 \pm 7.8$ |
| Selig *et al.* | $40 \pm 57$ | $1.7 \pm 6.4$ | $-2.6 \pm 9.9$ |
| Phase-Contrast Microscopy | | | |
| Proposed | $1 \pm 9$ | $-0.4 \pm 0.9$ | $0.0 \pm 0.9$ |
| Ruggeri *et al.* | $48 \pm 88$ | $3.8 \pm 3.2$ | $-0.4 \pm 1.9$ |

| Dataset | No. cells (%) | Mean cell area (%) |
|---|---|---|
| Fluorescence-Confocal Microscopy | | |
| Proposed at 1:1 | $0.5 \pm 1.4$ | $-0.3 \pm 1.4$ |
| Proposed at 1:25 | $0.3 \pm 1.0$ | $-0.1 \pm 1.0$ |
| Chiu *et al.* | $-1.7 \pm 1.6$ | $2.0 \pm 1.8$ |

phase-contrast microscopy dataset, the cell edges in the gold standard had a thickness of 2-3 pixels, so skeletonization was applied to reduce it to 1 pixel thickness.

A statistical analysis based on linear mixed-effects models [27] was performed to determine, for each parameter, whether there was a statistically significant difference in precision (smaller variance) and in accuracy (smaller absolute mean) between the two estimation errors. To determine whether the variances were different, we used a likelihood test to compare a model that assumes equal variances between both estimation errors with a model that assumes different variances. From the fixed effects test of the models we evaluated whether the mean values in both estimations were different. No correction for multiple testing was applied, and a p-value of $< 0.05$ was considered statistically significant.

*2) Specular Microscopy Dataset:* The statistical analysis indicated a significantly better precision in all parameters for the proposed method ($p < 0.001$, $p = 0.020$, and $p < 0.001$, for ECD, CV, and HEX respectively). The analysis also indicated a significantly better accuracy for CV ($p < 0.001$), but comparable accuracy for ECD and HEX ($p = 0.57$, both) (Table I).

*3) Confocal Microscopy Dataset:* This dataset was characterized for having significantly smaller images with a rather low cell density (Fig. 6d). The larger cell sizes (Fig. 7a-d) resulted in an initial oversegmentation in approximately 45% of the cells. In the output, there were on average only one over- and one undersegmented cell per image, the majority of them in the border region of the image. This provided a TPR of 97.5% and 0.016 FPs/cell. The statistical analysis indicated that, for ECD and CV, there was a significantly better precision ($p < 0.001$) and better accuracy ($p = 0.028$, $p = 0.002$, respectively) for the proposed method, but comparable precision and accuracy for HEX ($p = 0.074$, $p = 0.088$, respectively) (Table I). The small image sizes and large cells made the parameter estimation more sensitive to merging errors and small variation in edge location. Overall, the method could successfully cope with images exhibiting large differences in cell density (Fig. 6d).

*4) Phase-Contrast Microscopy Dataset:* This dataset showed a high cell density and a low degree of polymegethism (Fig. 6g-h), and the image quality was significantly better (Fig. 7e-h) compared to the other datasets. Therefore, only 7.5% of the cells were initially oversegmented. After applying our method, 20 of the images had a perfect merging, and the remaining 10 images showed on average one over- and one undersegmented cell. In total, the method provided a TPR of 98.70% and 0.0009 FPs/cell. The statistical analysis indicated a significantly better precision in all parameters for the proposed method ($p < 0.001$), a significantly better accuracy for ECD and CV ($p = 0.008$, $p < 0.001$, respectively), and similar accuracy for HEX ($p = 0.214$) (Table I). Overall, the high cell density and good quality image made that our method provided highly satisfactory estimates.

*5) Fluorescence Confocal Microscopy Dataset:* The last dataset was characterized by a large variation in cell size (Fig. 7i-l). Approximately 50% of the cells in each image were oversegmented, and around 5% of the cells were comprised of more than three fragments. In the output, there were on average three over- and three undersegmented cells (from the approx. 1000 cells per image). This provided a TPR of 99.37% and 0.0036 FPs/cell. However, the oversegmented cells were mainly large cells divided in 5-10 fragments (Fig. 7k-l), which affected the parameter estimation considerably. The statistical analysis indicated a significantly better accuracy in both parameters for our method ($p < 0.001$), but comparable precision ($p = 0.428$ and $p = 0.201$, for number of cells and mean cell area, respectively) (Table I). Due to the high intensity contrast between edges and cells, another SVM model with smaller cost (FP:FN of 1:25) could be added to the cascade, which barely introduced any FP but solved the majority of those oversegmented cells (figure not provided; values in Table I). Consequently, this also provided a statistically significant better precision in both parameters ($p = 0.024$ and $p = 0.007$, for number of cells and mean cell area, respectively).

## C. Computational Cost & Execution Time

Besides the details explained in Section II-A, several considerations were taken into account in order to reduce the computational cost. On average, each superpixel had 7 neighbors in the oversegmented image. Hence, the features of the superpixels were computed only once, stored in memory, and retrieved when building the feature vector of a candidate merger. When the features of a superpixel changed due to neighboring mergers, the features were recomputed and overwritten. Initially, the oversegmented image was translated into a structure where each superpixel, edge, vertex, and pixel was labeled, and where the connection between these elements were listed. This simplified the iterative merging process, as we could directly access to the elements that needed to be updated.

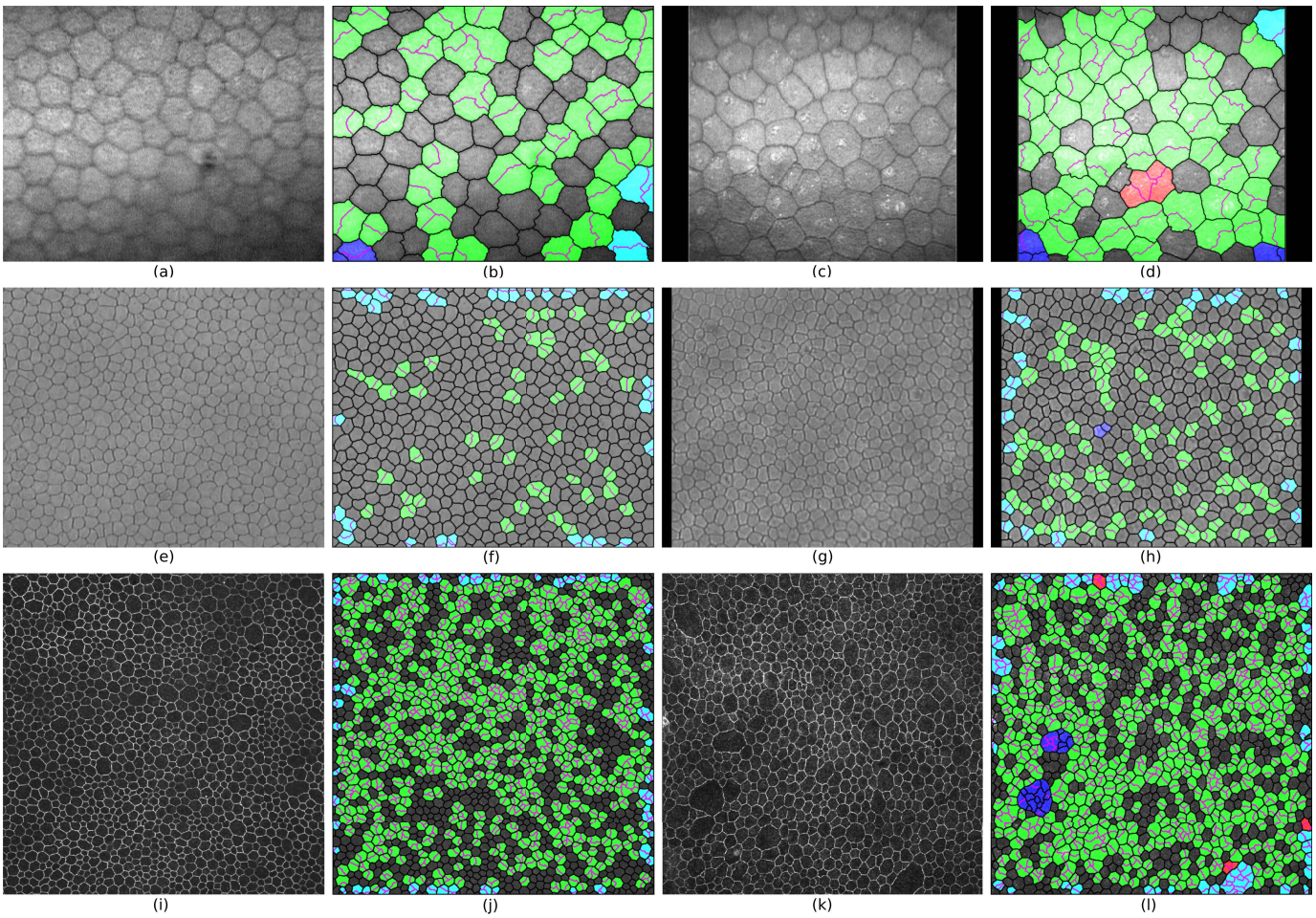



Fig. 7. Two representative images and their respective visual outcome of the merging for the confocal microscopy dataset of CE (a)-(d), phase-contrast microscopy dataset of CE (e)-(h), and fluorescence confocal microscopy dataset of RPE (i)-(l). Black lines indicate the segmentation outcome, and magenta lines indicate edges being removed in the merging process. All fragments involved in mergers are colored: green denotes complete cells after merging, red denotes undersegmented cells (FP), dark blue denotes oversegmented cell (FN), and light blue denotes results of mergers outside the gold standard area.

The computational cost of the 2-fragment classifier was linear with respect to the number ($N$) of initial superpixels, $\mathcal{O}(N)$. For the 3-fragment classifier, the relation was quadratic, $\mathcal{O}(N^2)$. For the main dataset, the entire segmentation process for a single pair of classifiers took on average 2.5 minutes (four times more if we used the cascade of four classifiers). The majority of time was used in evaluating all combinations of 3-fragments. Restricting the segmentation to mergers of two fragments reduced the computation time to less than 30 seconds, but as a consequence not all cells divided in 3 fragments were correctly segmented. We believe that these computation times can be decreased substantially if this method were implemented in a low-level programming language.

## IV. Discussion

We presented a new method to segment corneal endothelium images that requires no user intervention. The method provided very good results in endothelium images obtained from different devices and with large differences in cell density and variation in cell size. Furthermore, our method can be applied to other images of similar closely-packed cells with little extracellular matrix, such as RPE images. To the best of our knowledge, no other technique with such versatility has been reported up to now.

The qualitative results shown in Figs. 5 and 7 indicate accurate cell segmentation despite the presence of image artifacts and blurriness, and the variability in cell size was not a performance-limiting factor. The quantitative results indicate an average error of less than 2.1% (Table I), although this can be considerably smaller depending on the dataset and the estimated parameter. The proposed method outperforms the other automatic techniques from the literature, offering a significantly better accuracy and/or precision for the majority of the clinical parameters.

While we have used only classifiers dealing with cells divided in two and three fragments, it is straightforward to extend our method to cells composed of more fragments. Endothelial cells rarely show such a large difference in cell size that inclusion of explicit classifiers for cells divided in more than three fragments is required. Indeed, those cases only represent 0.35% of the cells in the specular microscopy

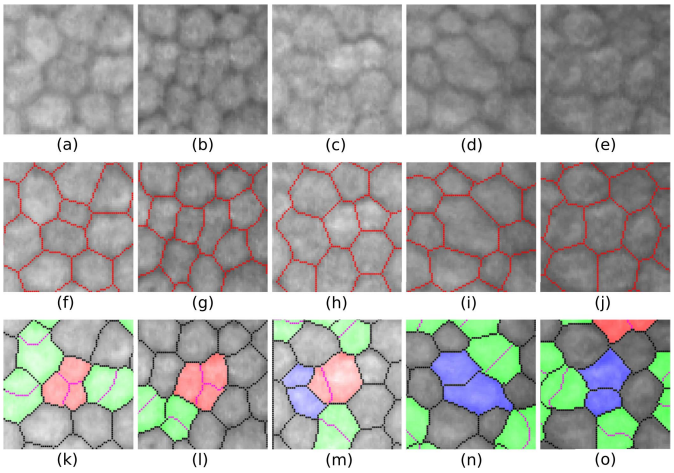



Fig. 8. Several examples of undersegmented cells (k)-(m), and oversegmented cells (m)-(o) in the segmentation result of the specular microscopy dataset, where black lines indicate the final segmentation and magenta lines indicate edges being removed in the merging process. All fragments involved in mergers are colored: green denotes complete cells after merging, red denotes undersegmented cells (FP), and dark blue denotes oversegmented cell (FN). Intensity image (a)-(e), gold standard in red (f)-(j).

dataset and 2.5% of the cells in the confocal microscopy dataset. In practice, the proposed method achieves the correct segmentation for the majority of those cells as the intensity-related features help to identify the false edges. This was illustrated convincingly in the fluorescence confocal microscopy dataset (RPE), where cells divided in up to 15 fragments were correctly segmented (Fig. 7i-l).

Among the datasets presented here, the specular microscopy images generate the highest number of errors due to their lower signal-to-noise ratio and lower image quality in terms of contrast and blurriness. Indeed, lack of contrast between edges and cells produces a rather inaccurate delineation of the edges in the initial oversegmentation, which is the major factor for errors in the merging process (Fig. 8a, 8c-e). Even in the absence of cell fragmentation, the inaccurate edge delineation significantly affects the estimation of CV and, especially, HEX. Other factors, such as oddly shaped cells (Fig. 8b, 8d) or cells touching in the image borders, may produce wrong mergers. We expect that the latter could be avoided with more training examples.

Overall, we observe that fragmented cells with correctly segmented edges are satisfactorily merged. Hence, to further increase the accuracy of the segmentation, improvements should be primarily made in the method to generate the initially oversegmented images.

In general, the method provides clinically usable results, especially for ECD and CV. As observed in Fig. 6, the error of those two biomarkers are relatively low, although the error for HEX is still rather large. Whether this accuracy is clinically acceptable depends more on the actual application and/or disease under study. Nonetheless, a manual correction could be suggested to correct mistakes. This could be done in a user-friendly manner, permitting the user to remove a false edge (or to undo an erroneously removed true edge) with a simple click, without the need of tracing it manually.

## Acknowledgment

The authors would like to thank Alfredo Ruggeri and Fabio Scarpa for providing us with the resulting segmentations of their algorithm [15], and Esma Islamaj, Caroline Jordaan, and Annemiek Krijnen for acquiring the images of our dataset.

## References

[1] G. O. Waring, W. M. Bourne, H. F. Edelhauser, and K. R. Kenyon, "The corneal endothelium: Normal and pathologic structure and function," *Ophthalmology*, vol. 89, no. 6, pp. 531–590, 1982.

[2] P. A. Mohammad-Salih, "Corneal endothelial cell density and morphology in normal Malay eyes," *Med. J. Malaysia*, vol. 66, no. 4, pp. 300–303, 2011.

[3] W. J. Armitage, A. D. Dick, and W. M. Bourne, "Predicting endothelial cell loss and long-term corneal graft survival," *Invest. Ophthalmol., Vis. Sci.*, vol. 44, no. 8, pp. 3326–3331, 2003.

[4] W. M. Bourne, "Biology of the corneal endothelium in health and disease," *Eye*, vol. 17, no. 8, pp. 912–918, 2003.

[5] I. Jalbert, F. Stapleton, E. Papas, D. F. Sweeney, and M. Coroneo, "*In vivo* confocal microscopy of the human cornea," *Brit. J. Ophthalmol.*, vol. 87, no. 2, pp. 225–236, 2003.

[6] A. S. Kitzmann, E. J. Winter, C. B. Nau, J. W. McLaren, D. O. Hodge, and W. M. Bourne, "Comparison of corneal endothelial cell images from a noncontact specular microscope and a scanning confocal microscope," *Cornea*, vol. 24, no. 8, pp. 980–984, 2005.

[7] M. L. Salvetat, M. Zeppieri, F. Miani, L. Parisi, M. Felletti, and P. Brusini, "Comparison between laser scanning *in vivo* confocal microscopy and noncontact specular microscopy in assessing corneal endothelial cell density and central corneal thickness," *Cornea*, vol. 30, no. 7, pp. 754–759, 2011.

[8] M. Hara, N. Morishige, T. Chikama, and T. Nishida, "Comparison of confocal biomicroscopy and noncontact specular microscopy for evaluation of the corneal endothelium," *Cornea*, vol. 22, no. 6, pp. 512–515, 2003.

[9] J. Huang, J. Maram, T. C. Tepelus, S. R. Sadda, V. Chopra, and O. L. Lee, "Comparison of non-contact specular and confocal microscopy for the evaluation of the corneal endothelium," *Invest. Ophthalmol., Vis. Sci.*, vol. 55, no. 13, p. 999, 2014.

[10] S. Jonuscheit, M. J. Doughty, and K. Ramaesh, "*In vivo* confocal microscopy of the corneal endothelium: Comparison of three morphometry methods after corneal transplantation," *Eye*, vol. 25, no. 9, pp. 1130–1137, 2011.

[11] M. O. Price, K. M. Fairchild, and F. W. Price, "Comparison of manual and automated endothelial cell density analysis in normal eyes and DSEK eyes," *Cornea*, vol. 32, no. 5, pp. 567–573, 2013.

[12] C. M. Klais, J. Buhren, and T. Kohnen, "Comparison of endothelial cell count using confocal and contact specular microscopy," *Ophthalmologica*, vol. 217, no. 2, pp. 99–103, 2003.

[13] C. Hirneiss, R. G. Schumann, M. Gruterich, U. C. Welge-Luessen, A. Kampik, and A. S. Neubauer, "Endothelial cell density in donor corneas: A comparison of automatic software programs with manual counting," *Cornea*, vol. 26, no. 1, pp. 80–83, 2007.

[14] B. Selig, K. A. Vermeer, B. Rieger, T. Hillenaar, and C. L. L. Hendriks, "Fully automatic evaluation of the corneal endothelium from *in vivo* confocal microscopy," *BMC Med. Imag.*, vol. 15, pp. 1–13, Apr. 2015.

[15] A. Ruggeri, F. Scarpa, M. De Luca, C. Meltendorf, and J. Schroeter, "A system for the automatic estimation of morphometric parameters of corneal endothelium in alizarine red-stained images," *Brit. J. Ophthalmol.*, vol. 94, no. 5, pp. 643–647, 2010.

[16] F. Scarpa and A. Ruggeri, "Development of a reliable automated algorithm for the morphometric analysis of human corneal endothelium," *Cornea*, vol. 35, no. 9, pp. 1222–1228, 2016.

[17] Y. Gavet and J. C. Pinoli, "Visual perception based automatic recognition of cell mosaics in human corneal endothelium microscopy images," *Image Anal. Stereol.*, vol. 27, no. 1, pp. 53–61, 2008.

[18] M. S. Sharif, R. Qahwaji, E. Shahamatnia, R. Alzubaidi, S. Ipson, and A. Brahma, "An efficient intelligent analysis system for confocal corneal endothelium images," *Comput. Methods Programs Biomed.*, vol. 122, no. 3, pp. 421–436, 2015.

[19] X. Song, L. Zhou, Z. Li, J. Chen, B. Yan, and L. Zeng, "Interactive image segmentation based on hierarchical superpixels initialization and region merging," in *Proc. Int. Congr. Image, Signal Process.*, 2014, pp. 410–414.

[20] L. Li, J. Yao, J. Tu, X. Lu, K. Li, and Y. Liu, "Edge-based split-and-merge superpixel segmentation," in *Proc. IEEE Int. Conf. Inf. Autom.*, Aug. 2015, pp. 970–975.

[21] C.-Y. Hsu and J.-J. Ding, "Efficient image segmentation algorithm using SLIC superpixels and boundary-focused region merging," in *Proc. 9th Int. Conf. Inf., Commun. Signal Process.*, 2013.

[22] S. Zhu, D. Cao, Y. Wu, and S. Jiang, "Improved accuracy of superpixel segmentation by region merging method," *Frontiers Optoelectron.*, vol. 9, no. 4, pp. 633–639, 2016.

[23] S. J. Chiu, C. A. Toth, C. B. Rickman, J. A. Izatt, and S. Farsiu, "Automatic segmentation of closed-contour features in ophthalmic images using graph theory and dynamic programming," *Biomed. Opt. Exp.*, vol. 3, no. 5, pp. 1127–1140, 2012.

[24] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Proc. Adv. Large Margin Classifiers*, 2000, pp. 61–74.

[25] D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: Modeling, analysis, and validation," *Med. Phys.*, vol. 31, no. 8, pp. 2313–3230, 2004.

[26] J. D. Ding *et al.*, "Anti-amyloid therapy protects against retinal pigmented epithelium damage and vision loss in a model of age-related macular degeneration," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 28, pp. E279–E287, 2011.

[27] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data* (Springer Series in Statistics). New York, NY, USA: Springer, 2000.