

Automated detection and classification of nuclei in immunohistochemical stainings for Fuchs' endothelial corneal dystrophy

Thomas Janssens^{1*}, An-Katrien De Roo^{2*}, Beatrijs Foets³, Joost J. van den Oord², Greet Van den Berghe¹, Fabian Güiza¹

¹ Laboratory for Intensive Care Medicine, KU Leuven

² Translational Cell & Tissue Research, KU Leuven

³ Department of Ophthalmology, UZ Leuven, Belgium

Abstract. Fuchs' endothelial corneal dystrophy (FECD) is a degenerative disease that affects the elderly population, and which lacks a unifying pathogenic theory and tangible drug targets. Immunohistochemical stainings can be used to identify proteins involved in the pathogenesis of FECD. We introduce a method for the automatic quantification of the ratio of stained cells starting from full high-resolution cornea images. First, the endothelium is extracted using entropy information in a low-resolution resampling. Then, within the endothelium, we heuristically detect and classify nuclei based on their size, color, and the color of the surrounding cytoplasm. This method achieves comparable results to manual evaluation in a set of corneas of patients with and without FECD.

1 Introduction

Fuchs' endothelial corneal dystrophy (FECD) is a progressive degenerative disease of the corneal endothelium (CE). The corneal endothelium is a thin (4 μm thick) monolayer of hexagonal cells that reside on Descemet's membrane, and that covers the inner layer of the transparent part of the eye. Normal CE keeps the cornea transparent through its selective barrier and active pump function. FECD is characterized by loss of CE cell density, resulting in corneal swelling and visual impairment. There are two forms of FECD, early- and late-onset FECD. This work focuses on late-onset FECD, which is the most common variant and affects approximately 4% of the population over the age of 40 in the United States[1]. FECD is the second most common indication for corneal transplantation performed in the elderly[2]. The prevalence of FECD and the need for donor corneas are expected to rise with aging of the population[3], hence the importance to identify molecular targets for drug therapy. So far, many factors have been found to contribute to the pathophysiology of FECD[4], such as oxidative stress[2], but a unifying theory and tangible drug targets are lacking.

Immunohistochemistry (IHC) is often used to study the characteristics of FECD CE compared to normal CE on a protein level. CE can display mosaic staining patterns[5], meaning that the result of a staining often is not uniformly positive or negative, but rather a proportion of corneal endothelial cells (CECs)

* These authors contributed equally to this work

will stain. The loss of CECs in FECD causes the remaining cells to stretch over the gaps that are created, so that the stained surface area doesn't correlate directly with the absolute number of cells which are positive, nor does the absolute positive cell count correlate with the ratio of positive versus negative cells. We therefore use the ratio of positive CECs divided by the total number of CECs to quantitate and compare IHC stainings.

Counting of positive CECs is traditionally done in an analogous way by manually counting cells while looking through a microscope. Automated digital image analysis can be an alternative to obtain more objective, robust and reproducible results. FRIDA Software [6] has been used to study CE of patients with FECD, but it requires manual region selection and only produces a pixel area quantification, not individual nuclei detection. This makes it impossible to compare nuclei ratios. In this work we present a novel automatic method to quantify individual nuclei in IHC stainings of CE specimens in an accurate and reproducible manner on high-resolution light microscopy slide images, taking into account distinctive characteristics of FECD, and without need for a manual area of interest selection.

The extreme scale discrepancy between image and cell presents a challenge. The images are high-resolution composites comprising the entire tissue slide, measuring in excess of 20.000*30.000 pixels. In contrast, the desired nuclei are exclusively present on a thin (around 20 px wide) strip of CE, and have a surface area of around 5.000 px. This translates to a 10^5 difference in scale, making a multi-scale approach essential for an optimal solution, where we select the correct scale at each step to balance accuracy and performance. The cornea needs to be separated from artifacts, such as mounting medium or marks on the slide, which can be in close proximity (visible in Fig. 4).

First, we detect the cornea region in a low resolution resampling of the slide. Next, we identify the CE layer in the cornea based on a medium resolution resampling. Finally, we examine the CE at full resolution to identify the nuclei and their cytoplasm's staining (overview in Fig. 1). An adaptable algorithm is used, extensible enough to deal with different types of staining. To evaluate our method, we compare it against an expert manual quantification, and show that it provides a good approximation of the staining ratio measure.

2 Methods

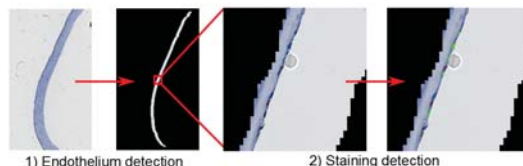


Fig. 1: Pipeline steps: Endothelium and staining detection

2.1 Image acquisition and processing

Tissue slides were scanned with the NanoZoomer 2.0-RS (Hamamatsu, Hamamatsu City, Japan) at 400x magnification. The average corneal diameter is about 11.5 mm, while the transplanted full-thickness corneas for FECD measure 7 mm in diameter. Because the staining pattern of central and peripheral corneal endothelium could differ, rectangles with a height of 7 mm were manually superposed onto the central corneas in the scanned images to provide equivalent areas for comparison. Our algorithm works on the full-resolution cornea as delineated by these rectangles.

2.2 Endothelium detection

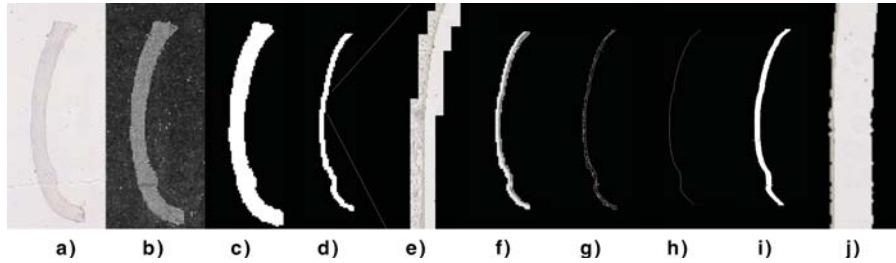


Fig. 2: Endothelium detection: a) Original image I , b) entropy image E_{s_1} , c) mask M^D , d) CE site $M_{s_1}^{E_n}$, e) overlay (detail), f) entropy image E_{s_2} , g) Sobel edge, h) concave edge part, i) expanded edge, j) final endothelium detection (detail).

Since the corneas can be presented in both horizontal and vertical orientation, our first step is to rotate all horizontal corneas (width of the bounding box exceeding height) 90° clockwise, thus simplifying subsequent steps.

We work with the assumption of a slightly curved shape for the cornea, where the endothelium is present at the most concave side. We also assume this cornea lies relatively free of other tissue and artifacts such as droplets of mounting medium or other artifacts. In order to find the epithelium we need to a) determine the location of the concave side of the cornea (left or right) b) extract the epithelial layer present at this concave side as precisely as possible.

In order to detect the boundaries of the cornea against the background in a robust and computationally efficient way, we transform the original image $I \in R^{M \times N \times 3}$ (Fig. 2a) into a downscaled red channel entropy image E_{s_1} with scaling parameter s_1 , where $E_{s_1}(x, y) = E(I_{s_1 x \dots s_1(x+1), s_1 y \dots s_1(y+1)}^R)$ for $0 \leq x < M/s_1, 0 \leq y < N/s_1$, and $E(I)$ representing the entropy of an 8-bit grayscale image. s_1 is chosen so that one pixel in E_{s_1} covers more than the expected width of the endothelium in I (Fig. 2b).

We have opted for an entropy image because although the cornea itself shows a variety in structures, there is dependably more structural variation (i.e. entropy) in any layer compared to the blank background, meaning E_{s_1} consists of

a more or less uniformly dark background and a bright cornea. In order to differentiate these two we apply a threshold determined using Otsu's method [7], creating a binary image mask M^O , where $\forall(x, y) \in M^O : (x, y) \in R^{\{0,1\}} \wedge (x, y) = 1 \Leftrightarrow E_{s_1}(x, y) > O(E_{s_1})$, where $O(G)$ is a function that finds an Otsu threshold for a grayscale image G .

Because of the possible proximity of image artifacts that are not part of the cornea itself, we perform a morphological erosion to move the boundaries slightly inward; $M^E = M^O \ominus S$, with S a 2px square ($\approx 1 \mu\text{m}$ on the slide).

If M^E contains multiple 4-connected regions, we retain only the largest one as the cornea is several times bigger than any surrounding artifacts, creating a filtered mask M^F . This region is dilated again: $M^D = M^F \oplus S$, which in effect results in a morphological opening on this largest region.

This leaves us with a low-resolution binary mask M^D showing the cornea location (Fig. 2c). The next step is to determine the direction of its orientation, i.e. determining the concave side housing the CE. For this we use a relatively simple heuristic. First, M^D is cropped so that the cornea region touches all sides of the image. Next, 10 points on the edge of the cornea are measured: the 5 leftmost and 5 rightmost points of the cornea at image row 1, $\lfloor 1/4H \rfloor$, $\lfloor 1/2H \rfloor$, $\lfloor 3/4H \rfloor$, H , where H is the number of rows in the cropped image.

We define the convexity C_L of left side by the horizontal distance of the left non-middle points towards the left edge of the image compared to the center point. This is done in a similar fashion for the right side (C_R). If $C_L > C_R$, the left side is taken to be the convex side, otherwise we take the right side. Though simple, this heuristic proved to be enough to reliably detect the orientation of all corneas satisfying the assumptions stated above.

Once the concave side is determined, we obtain the pixels $M_{s_1}^{En}$ containing the endothelium by choosing the left- or rightmost pixel in each image column (Fig. 2d,e).

As a further refinement, we remove the parts before and after the places where the cornea bends back towards the concavity, since these regions correspond to the transition from endothelium to cornea cross-section.

Once we have determined the endothelium site $M_{s_1}^{En}$ we extract the region covered by the selected pixels from the original image; each pixel in $M_{s_1}^{En}$ corresponds to a s_1 -width pixel block in I .

Since $M_{s_1}^{En}$ is still only an approximation of the true full-resolution endothelium location, we refine this result further. We again downsample I into an entropy image E_{s_2} (Fig. 2f), this time with a smaller scale factor ($s_2 = 25$). This is feasible because now we only examine the narrow strip of probable endothelium instead of the entire cornea image.

We then use Sobel edge detection (Fig. 2g) to find the endothelium location $M_{s_2}^{En}$, by taking the concave side using the previously detected orientation (Fig. 2h). This edge is then again upscaled to s_2^2 blocks in I using the same process as before. We pad the upscaled pixel block with a fixed distance (150 px / 37 μm inwards, 600 px / 150 μm outwards) empirically set to be sure to include the full

endothelium, but unlikely to contain stromal nuclei or artifacts (Fig. 2i,j, Fig. 4 shows the stroma removed).

2.3 Staining detection

Extracting the endothelium from the cornea image reduces the area of interest enough to do per-pixel processing to detect the nuclei. Since the nuclei are stained a dark blue with hematoxylin counterstaining, we can extract their pixels through a threshold on the hue, saturation and brightness channels: $\forall p : p \in N \Rightarrow RGB(p) \in \{[H_{min}, H_{max}], [S_{min}, S_{max}], [B_{min}, B_{max}]\}$, where p is a pixel and N the set of all pixels belonging to a nucleus. After applying this threshold we obtain a binary image. A connected-components analysis (8-way) results in a set of candidate nucleus regions. Here we impose thresholds to remove regions whose area is too small or large to be an individual nucleus: $R \subset N \Rightarrow A_{min} < size(R) < A_{max}$. A_{min} is 150 px², or about 9 μm², A_{max} is 1000 px², or about 62 μm².

We create a dilated copy R_D of each region R , dilating R by $d = 15$ pixels, or 4 μm. Next we subject all pixels in R_D to a HSB empirically set to detect the specific type of cytoplasmatic staining looked for, indicating a positive CEC. If the number of cytoplasmatically stained pixels in the area surrounding a nucleus passes a threshold of $\alpha = 30$, the nucleus as defined by R is considered to be part of a positively stained CEC. Nuclei below this threshold are marked negative.

Parameters A_{min} , A_{max} , d , α were empirically determined such that they would be valid for all evaluated stainings. To correctly detect each type of nuclear and cytoplasmic staining the HSB thresholds reported in Table 1 were used.

3 Experiments

In order to evaluate the performance of our method we compare its results against those of manual evaluation by an expert over a dataset consisting of several cornea images.

3.1 Acquisition of images

Corneas from patients with and without FECD were selected from the archives of the Department of Pathology, UZ Leuven. We chose three types of cytoplasmic IHC to test the algorithm, based on significantly differentially expressed genes in a microarray expression analysis (MEA) comparing CE from 4 patients with and 4 patients without FECD [data not shown]: brain-derived neurotrophic factor (BDNF), which is related to neural crest origin of CE; keratin 7 (KRT7), a cytoskeletal protein; and superoxide dismutase 3 (SOD3), which is related to oxidative stress. Corresponding antisera that were used are: BDNF rabbit polyclonal antibody (Santa Cruz Biotechnology, California, USA), Cytokeratin 7 monoclonal mouse antibody (Dako, Glostrup, Denmark), and anti-Superoxide Dismutase 3 mouse monoclonal antibody (Ab Frontier, Seoul, South Korea). Sections were counterstained with hematoxylin. BDNF and SOD3 stainings were done

manually, whereas KRT7 stainings were done by an autostainer, giving slightly different intensities in counterstaining. For each type of staining 6 corneas with FECD and 3 control corneas were evaluated for a total of 27 images. Corneas with distorted shape or large artifacts were not considered as they fall out of the scope of our method.

	H_{nuc}	S_{nuc}	B_{nuc}	H_{cyto}	S_{cyto}	B_{cyto}
KRT7	[131, 217]	[20, 255]	[0, 185]	$[0, 15] \cup [217, 255]$	[40, 255]	[0, 200]
SOD3	[120, 180]	[113, 255]	[0, 255]	$[0, 40] \cup [200, 255]$	[0, 15]	[0, 255]
BDNF	[128, 187]	[107, 255]	[0, 152]	$[0, 58] \cup [173, 255]$	[40, 255]	[0, 157]

Table 1: HSB threshold values for three used cytoplasmatic stainings, with each channel $[0, 255]$.

As a manual evaluation, a domain expert (second co-author) examined this dataset to provide a count of stained and unstained nuclei. We processed all images with the algorithm as well, first extracting the endothelium, then detecting the stained and unstained CEC counts. We use the stained vs. unstained ratio (R_{SU}) for all images as a measure to quantitate and compare IHC stainings.

Table 2 shows the median R_{SU} for manual and automatic evaluation of all data subsets. Table 3 shows two-sample Wilcoxon test p-values expressing the difference between automatic and manual evaluations, and FECD and control subsets. Fig. 3a shows the results of our model as compared to the expert’s evaluation. The correlation coefficient between the manual and automatic evaluation of R_{SU} for all 27 images is $r = 0.75$. For KRT7, SOD3 and BDNF separately, it is 0.72, 0.76, and 0.89, respectively. Evaluation box plots are shown in Fig. 3b.

Comparison	KRT7	SOD3	BDNF	KRT7 (c)	SOD3 (c)	BDNF (c)
Manual	1.22	0.64	0.45	0.05	0.05	0.55
Automatic	1.25	0.34	0.77	0.02	0.07	0.14

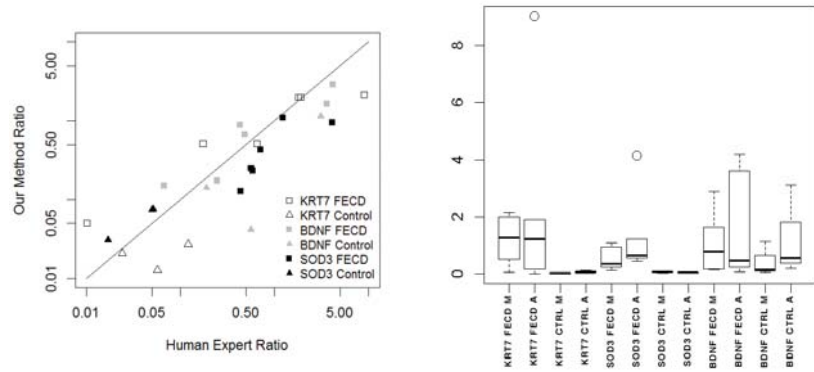
Table 2: R_{SU} medians for manual and automatic FECD and control corneas per staining type.

Comparison	KRT7	SOD3	BDNF
Manual FECD vs. Controls	0.17	0.02	1.00
Automatic FECD vs. Controls	0.02	0.02	0.26
Manual vs. Automatic FECD (paired)	0.84	0.03	0.69
Manual vs. Automatic Controls (paired)	0.22	0.25	0.25

Table 3: Two-sample Wilcoxon test p-values for manually vs. automatically obtained ratios, and FECD (3×6 samples) vs. controls (3×3 samples).

4 Discussion

As tables 2, 3 and Figs 3a and 3b show, our method approximates the expert ratios well, with the median ratio difference being much higher between the images of either method than the disagreement of the methods on individual images. In particular, the separation of FECD and controls in KRT7 and SOD3 of the manual evaluation is also present in the automatic evaluation. In future work, we will apply our method on a larger sample size and additional types of stainings to validate MEA data on a protein level.



(a) Manually versus automatically de- (b) Box plots of manual (M) and auto-
 tected ratios. matic (A) evaluation.

Fig. 3

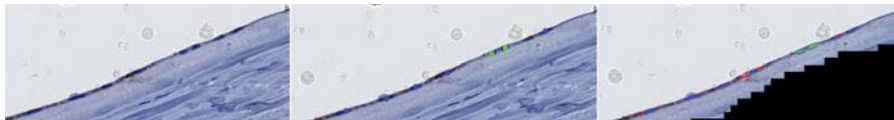


Fig. 4: Evaluation of a BDNF-stained CEC fragment. Left to right: base image, expert evaluation and evaluation by our method. Green denotes an unstained CEC, blue a stained CEC, red indicates cytoplasmic staining (only annotated for our method). Also note the visible drops of mounting medium above the cornea which are correctly discarded as artifacts by our method.

A limitation of our algorithm is that the absolute nuclei counts are not always precisely accurate. This is in part due to inaccuracies in the CE detection phase, where the detected layer might drift inwards, including stromal nuclei, or outwards, excluding some CE nuclei. Depending on calibration nuclei might also be missed, when they are out of focus or cut tangentially. Though the influence on the R_{SU} is diminished by the fact that it in effect only reduces the sampled area and doesn't favor either type of CEC.

The expert and the algorithm might disagree on the interpretation of colors, partly due to inaccuracies in the model, also because digital images are more difficult to interpret by an expert. Digitally, staining intensity and coloration can be difficult to determine for nuclei, making manual evaluation inevitably subjective. In that way, the automated analysis provides a more objective measure. Fig. 4 compares manual and automatic interpretations.

The multi-scale approach reduces the evaluation time greatly: in our method processing an already detected endothelium (750*30.000 px) takes around 3 minutes. Were we to apply this method to the entire (20.000*30.000 px) image, it would take 80 minutes, not including time taken to differentiate CE nuclei from other corneal nuclei. Manual evaluation of a cornea takes about 15 minutes if done digitally, and 5 minutes under a microscope.

Our method uses some heuristics with fixed empirically tuned parameters. Though this is robust enough for regularly-shaped corneas, it would be interesting to investigate more flexible methods to extend our method to correct for possible artifacts such as a warped cornea shape or irrelevant elements.

We have limited the scope of this work to cytoplasmatic stainings. It would be interesting to see how well it could be adapted to accommodate nuclear stainings, which are also of scientific interest. Besides R_{SU} , it would also be interesting to quantify and compare staining intensity between FECD and controls. Though this method focuses exclusively on the endothelium, both the stroma and the epithelium are also interesting targets for analysis, and could be isolated with minor changes to our method. The methodology presented here starts from a manually defined rectangular cornea image (Fig. 2a). In order to further reduce evaluation time in future work we will automate this first detection step.

5 Conclusion

We have created an algorithm capable of detecting and classifying endothelial cells in corneas present on a high-resolution full slide image, under a variety of stainings, accelerated by a multi-scale approach. This automatic method has thus been found valid to be used for comparison of immunohistochemical staining patterns of CEC between FECD and control corneas in future research projects.

6 Acknowledgments

An-Katrien De Roo is a PhD fellow of the Research Foundation, Flanders (FWO). The authors would also like to thank the Fund for Research in Ophthalmology (FRO, Belgium) and Private initiative Mieke Perdaens Fund for Eye Research, for supporting this research.

References

1. Krachmer, J.H., Purcell, J.J., Young, C.W., Bucher, K.D.: Corneal endothelial dystrophy: a study of 64 families. *Archives of ophthalmology* **96**(11) (1978) 2036–2039
2. Elhali, H., Azizi, B., Jurkunas, U.V.: Fuchs endothelial corneal dystrophy. *The ocular surface* **8**(4) (2010) 173–184
3. Jurkunas, U.V., Bitar, M.S., Funaki, T., Azizi, B.: Evidence of oxidative stress in the pathogenesis of fuchs endothelial corneal dystrophy. *The American journal of pathology* **177**(5) (2010) 2278–2289
4. Wright, A.F., Dhillon, B.: Major progress in fuchs’s corneal dystrophy. *The New England journal of medicine* **363**(11) (2010) 1072
5. Foets, B., Van den Oord, J., Volpes, R., Missotten, L.: In situ immunohistochemical analysis of cell adhesion molecules on human corneal endothelial cells. *British journal of ophthalmology* **76**(4) (1992) 205–209
6. : Framework for image dataset analysis. The Johns Hopkins University, <http://bui3.win.ad.jhu.edu/frida/>, 2007.
7. Otsu, N.: A threshold selection method from gray-level histograms. *Automatica* **11**(285-296) (1975) 23–27