AN AUTOMATED DRUSEN DETECTION SYSTEM FOR CLASSIFYING AGE-RELATED MACULAR DEGENERATION WITH COLOR FUNDUS PHOTOGRAPHS

Yuanjie Zheng¹, Brian Vanderbeek², Ebenezer Daniel², Dwight Stambolian², Maureen Maguire², David Brainard³, James Gee¹

¹Department of Radiology, ²Department of Ophthalmology, and ³Department of Psychology at the University of Pennsylvania, Philadelphia

ABSTRACT
We present a system of automated drusen detection from color fundus photographs with our ultimate goal being to automatically assess the risk for the development of Age-related Macular Degeneration (AMD). Our system incorporates learning based drusen detection and includes fundus image analysis techniques for image denoising, illumination correction and color transfer. In contrast to previous work, we incorporate both optimal color descriptors and robust multiscale local image descriptors in our drusen detection process. Our system was evaluated with color fundus photographs from two AMD clinical studies [1, 2]. By comparing our results to those obtained via manual drusen segmentation, we show that our system outperforms two state-of-the-art techniques.

1. INTRODUCTION
Age-related Macular Degeneration (AMD) is the most common cause of blindness in developed world [1, 3]. The early stage of AMD is an ocular condition associated with minimal visual impairment and characterized by drusen and pigmen-
tary abnormalities in the macula, and degeneration of the reti-
nal pigment epithelium (RPE) [1]. The late stage of AMD is characterized by geographic atrophy (GA), RPE detachment, choroidal neovascularization (CNV) and disciform scar. Only the late stage of AMD results in moderate and severe losses in visual function. However, the presence of drusen is a hallmark of AMD, and magnification/multiplication of drusen indicates increased risk of eventual visual loss from AMD.

Grading drusen in color fundus photographs (CFPs) is an important component in the established classifications of AMD [3, 1]. Traditional grading methods [3, 1] used to quantify drusen are based on the use of overlay standard circles and subjective evaluation. The grader is asked to pick one from several standard circles to match each drusen in order to decide its size, to mentally aggregate the total drusen area, and to convert the net result into a categorical number. Although important relationships between these manual grading values and AMD progression, these methods are labor-intensive, difficult to reproduce, and may lose important information.

Automated drusen detection with computerized algorithms [4, 5, 6] is potentially capable of making grading of drusen more rigorous, reproducible, quantitative and cost-effective. However, digital techniques have not as of yet gained widespread acceptance for several reasons (as shown in Fig. 1). First, it can be challenging to reliably localize drusen against the varying background of the pigments of the macula, RPE and choroid, and to differentiate drusen from areas of RPE hypopigmentation, exudates and scars. Second, the inherent nature of the reflectance of even a normal macula is nonuniform, and this presents an obstacle for automatic drusen segmentation. Third, color variations between subjects in CFPs, caused by large natural variations in choroidal pigmentation and iris color, can mask the more subtle varia-
tion between drusen and other lesions or background.

In this paper, we present an automated color-fundus drusen detection system with our ultimate goal being to automatically assess the risk for development of AMD. Our system combines a set of computerized algorithms for pattern recognition, computer vision and machine learning. It is robust and accurate (as shown in Fig. 1) in the face of challenges described above, as validated with CFPs from two milestone AMD clinical studies [1, 2].

2. SYSTEM OF AUTOMATED DRUSEN DETECTION
2.1. System Overview
As illustrated in Fig. 2, the integrated image analytics of our system consist of several CFP preprocessing procedures. These include image denoising (with the Non-Local Mean Filter in [7]), retina mask generation (simply by image thresholding and hole filling with certain morphological operations), illumination correction to correct the slow background variation with our technique described in [8], and color transfer to render all testing images similar in color. Our color transfer is accomplished in a group-wise way by first creating an average intensity histogram over all testing CFPs, and then transforming the intensity of each CFP so that the histogram of the output approximately matches this average histogram. These two procedures are repeated until convergence. From the preprocessed CFPs, drusen are then detected using a learning based scheme, which will be detailed in the following section.
2.2. Learning Based Drusen Detection

Our drusen detection scheme relies on machine learning techniques and consists of two sequential procedures of a pixel-wise classification and a region-wise classification (as shown in Fig. 2). The pixel-wise classification independently determines whether each pixel is a drusen or not using image features obtained from a series of robust local appearance descriptors. It is accomplished with a powerful Ada-LS-SVM classifier, i.e. exploiting AdaBoost [9] for feature selection followed by LS-SVM [10] for classification. The region-wise classification, introduced in order to remove false-positives from the pixel-wise classification, applies a set of region based features to estimate whether spatially connected components of output of the pixel-wise classification are a drusen or not. Region-wise classification is carried out with the LS-SVM directly.

2.2.1. Features in Pixel-wise Classification

Image features used by the pixel-wise classification in our system are obtained from a set of optimal image color descriptors and a series of multiscale image local descriptors. The color descriptors characterize the local photometric properties of image and include the hue histogram and color moment invariants in [11]. They are computed within a local image patch in an optimal extent specified by an optimal image scale, which in turn is obtained with the method in our previous work [12].

The multiscale image local descriptors are defined to characterize the local geometric structure of the image and are motivated primarily by two observations. First, drusen in CFP at a specific scale is geometrically different from many of other retinal components, as shown in Fig. 3 (e)-(h). Second, features characterizing these geometrical attributes demonstrate very distinct variation patterns across different image scales (as shown in Fig. 3 (c)(d)) and these patterns carry abundant information useful for drusen detection.

We employ two types of features to characterize these multiscale image properties: the Hessian features previously proposed by us in [12] and a series of novel TV (total variation) features. Hessian features [12] at an arbitrary pixel located at \((x,y)\) can be denoted by \(\Lambda(x,y) = \{\lambda_1(x,y,\sigma), \lambda_2(x,y,\sigma); \sigma \in \Sigma(x,y)\}\) where \(\lambda_1\) and \(\lambda_2\) represent two eigenvalues of the image Hessian matrix at scale \(\sigma\), and where \(\Sigma(x,y)\) is a sequence of optimal image scales at \((x,y)\) (as detailed in [12]).

Our TV features are obtained by solving the following TV model [13] based minimization with a sequence of different values of the adjusting parameter \(\gamma\).

\[
\min_{u \in BV(\Omega)} |u|_{BV(\Omega)} + \gamma |I - u|_{L^1(\Omega)}
\]

where \(I\) denotes the green channel of the given CFP, \(\Omega \subseteq \mathbb{R}^2\), and where the definitions of the BV (bounded variation) space and \(L^1\) norm space are provided in [13]. Eq. (1) generates a smoothed version \(u\) of the given image in an anisotropically-filtering way which can remove image oscillations while p-reserving image edges. As pointed out in [13], \(\gamma\) in Eq. (1) indicates the scale of image analysis. The resulting value \(u\) at an arbitrary pixel is determined by the similarity of \(I\) between its neighborhoods in an extent specified by \(\gamma\).

Similar to the Heissian features used in [12], we specify
Fig. 3. Multiscale image local descriptors employed in our drusen detection system. (a): Given fundus image. (b): Given image overlayed with 5 rectangles with colors of red, green, blue, cyan, and black, respectively, which enclose a large drusen, a small drusen, a bright region of normal retina, a dark region of normal retina, and a section of vessel, respectively. (c): Multiscale Hessian matrix eigenvalues and (d): Multiscale TV filtered image values of the locations specified by the center of the color-matched rectangle in (b). (e)-(h): plot of 3-D surfaces for the red, green, blue, and black rectangles in (b), respectively. (i)-(l): TV filtered images using Eq. (1) with $\gamma = 0.2, 0.4, 0.8, 2$, respectively.

We choose the smallest scale in $S_{\gamma}(x,y)$ as the optimal scale of the local geometrical structure and denote it as $\gamma^*(x,y)$.

Then, we specify a series of $\gamma$ values for $(x,y)$, denoted and computed by $\Gamma(x,y) = \{\gamma^*(x,y) - M_{\gamma}, \delta\gamma, \cdots, \gamma^*(x,y) - \delta\gamma, \gamma^*(x,y), \gamma^*(x,y) + \delta\gamma, \cdots, \gamma^*(x,y) + M_{\gamma}\delta\gamma\}$, where $\delta\gamma$ (e.g. 0.2) denotes the step size and $M_{\gamma}$ (e.g. 4) determines the number of scales. Therefore, we finally obtain a set of TV features: $U(x,y) = \{u(x,y,\gamma) : \gamma \in \Gamma(x,y)\}$ to represent the local geometrical structure of $(x,y)$.

2.2.2. Features in Region-wise Classification

Region-level features are used to characterize the morphological or photometric properties of the candidate drusen regions (similar to [4]), including region area, region average values of the green and red channels, regional maximum brightness, border average brightness value, border maximum brightness, border minimum brightness, border average brightness gradient, and border brightness gradient’s standard deviation.

3. RESULTS

The proposed drusen detection system was validated by fundus photographs from the CAPT study [1] and Amish study [2]. The CAPT set consists of 50 stereoscopic CFPs of the macula, digitized from color film captured with a Zeiss 300 fundus camera. A trained rater (who is not an ophthalmologist) manually segmented drusen areas. The Amish set contains 88 CFPs obtained with a Canon EOS-D10 digital camera and by screening Amish individuals (for AMD) [2] over the age of 50 years living in Lancaster County, Pennsylvania, USA. A trained rater (who is not an ophthalmologist) also manually marked the drusen areas. In addition, an ophthalmologist (Brian Vanderbeek who is the second author of this paper) extensively studied these retinal images and manually segmented drusen for 14 representative images.

Validation of our system was made by contrasting its output with the manually marked drusen areas on a rigorous pixel-by-pixel basis and statistically comparing with two state-of-the-art drusen detection algorithms: STARE [4] and HALT [5]. For the learning based algorithms, our system and STARE, a leave-one-out strategy was adopted. We plotted the receiver operating characteristics (ROC) curves, computed the area under curve (AUC) values, and obtained the values of accuracy, sensitivity and specificity. From the results in

1We exploited our own implementations of STARE and HALT (no free software tools available for these two algorithms).
Table 1. Statistical results (Accu: accuracy; Sens: sensitivity; Spec: specificity) of our system, STARE [4] and HALT [5] in drusen detection on CAPT CFPs with drusen delineated by a trained rater - denoted by “CAPT(Rater)”, Amish CFPs with drusen delineated by an ophthalmologist - denoted by “Amish(Ophth)”, and Amish CFPs with drusen delineated by a trained rater - denoted by “Amish(Rater)”.}

<table>
<thead>
<tr>
<th>Data set (operator)</th>
<th>CAPT(Rater)</th>
<th>Amish(Ophth)</th>
<th>Amish(Rater)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accu</td>
<td>0.80</td>
<td>0.86</td>
<td>0.83</td>
</tr>
<tr>
<td>Sens</td>
<td>0.82</td>
<td>0.87</td>
<td>0.85</td>
</tr>
<tr>
<td>Spec</td>
<td>0.75</td>
<td>0.78</td>
<td>0.71</td>
</tr>
<tr>
<td>Accu</td>
<td>0.86</td>
<td>0.80</td>
<td>0.77</td>
</tr>
<tr>
<td>Sens</td>
<td>0.87</td>
<td>0.82</td>
<td>0.80</td>
</tr>
<tr>
<td>Spec</td>
<td>0.78</td>
<td>0.67</td>
<td>0.60</td>
</tr>
<tr>
<td>Accu</td>
<td>0.80</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td>Sens</td>
<td>0.82</td>
<td>0.82</td>
<td>0.81</td>
</tr>
<tr>
<td>Spec</td>
<td>0.68</td>
<td>0.65</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Fig. 4. ROC curves of our system, STARE [4] and HALT [5] in drusen detection on CAPT CFPs with drusen delineated by a trained rater - denoted by “CAPT(Rater)”, Amish CFPs with drusen delineated by an ophthalmologist - denoted by “Amish(Ophth)”, and Amish CFPs with drusen delineated by a trained rater - denoted by “Amish(Rater)”.

Fig. 4 and in Table 1, we found our system outperforms state-of-the-art techniques. In Fig. 1, we demonstrate the drusen areas found by our system for one CAPT CFP and one Amish CFP. We also found that classification improvement with the pixel-wise features selected by AdaBoost compared with the unselected features is significant ($p < 0.005$).

4. CONCLUSION AND FUTURE WORK

We present an automated drusen detection system with our ultimate goal of classifying age-related macular degeneration (AMD) from color fundus photographs (CFPs). It consists of several image preprocessing procedures and a learning based drusen segmentation scheme which integrates both optimal image color descriptors and a sequence of robust multiscale local image descriptors. As validated using our manual drusen delineations, our system outperforms two state-of-the-art algorithms. It can be a clinically invaluable tool for the standardization of drusen evaluation in AMD classification.

Future work includes studies on the correlation between our system and the manual grading scales for drusen related AMD characteristics.

5. REFERENCES