Analysis of reviewers' comments

Reviewer A

Reviewer's Comment	Authors' Comment	Proposed Action	Done
This paper presents a sequence of image processing operations for segmenting out blood vessels from images of a retina. In short, Gabor-type filters are used as a preprocessing stage, followed by a global thresholding stage based on a previously-published threshold selection procedure.	The main focus of the paper is to draw the reader's attention to the effect of gabor filter normalization on automatic segmentation based on thresholding. This is an issue that is overlooked in previous literature. To illustrate this effect, a review of existing gabor normalization techniques has been carried out and contributions have been made for improvement.	N/A	N/A
The method is evaluated on a standard dataset and the performance is good. However, in my opinion the methods proposed are simplistic, and aside from some minor variations in algorithmic details there is minimal exploration around possible variations. There is no indication that alternative methods were investigated before the proposed ones where found, and I find it hard to believe that the problem is so easy.	An extensive literature review was conducted. The divisive and zero-integral methods are derived from this literature. The offset method was conceptualized by a realization that dimensionality reduction and preservation of neighbourhood interpixel relations are fundamental to effective vessel enhancement. Results are compared to those presented in the literature.	N/A	N/A
The paper is also inaccurate in many places and grossly incomplete in others. For example, while the Gabor kernel has many degrees of freedom the only clue as to how the filter responses are used in the subsequent thresholding is this: "When each pixel is filtered, all wavelengths within the range are considered and the maxumum response is recorded". Within what range? Since wavelength is a real number some discretisation must have been performed. How? Similarly, Gabor kernels have a real-valued rotation parameter that is completely omitted in any discussion of how the results are generated. A much higher level of precision is required in a scientific publication.	The wavelengths that were used are 4, 8 and 12. These are shown in figure 4. The orientations were rotated by 15 degrees in the range [0, 180) degrees as is the case in previous literature. Assuming all 3 wavelengths are used, each combination of wavelength and orientation is considered, resulting in 12X3 combinations. Each of these wavelength-orientation combinations is used to create a Gabor kernel of size 21X21. Every pixel that is within the FOV is convolved with each of these 36 kernels. This results in 36 Gabor filter responses. The highest out of all the responses is chosen. The response image constitutes of the highest gabor filter response for each pixel. This image is then thresholded. Thresholding is a process of selecting a graylevel to be used as a benchmark for deciding whether a pixel should be white or black. Typically values below the threshold are set to black and those equal or above are set to white. Thresholding can be done manually or automatically. Both	Will state the exact wavelengths and orientations used and motivate in section 2.	✔

	approaches are investigated.		
I'm surprised at the apparent quality of the results, especially given that the presentation seems not to display the level of critical insight I would expect to see from a method that is apparently close to state of the art. If the authors made an easily runnable version of their code available which could be tested on the relevant datasets, then the results might be more compelling.	The methods proposed are by no means the best as far as retinal segmentation is concerned. Literature with higher accuracies has been cited. The argument that is put forward is that gabor filter normalization can affect the quality of segmentation. This is what the results demonstrate.	Will make a demo application and its code available to reviewers.	✔
p.1: "gabor" should be "Gabor" (throughout) p.1: "zero-Integral" should be "zero-integral" p.1: "being sort after" should be "being saught after" p.1: "[8].The" should be "[8]. The" p.1: "comprises of methods" should be "comprises methods" p.1: "Machines(SVM)" should be "Machines (SVM)"		Will do	✔
p.2: "vessels are generally piece-wise linear" sounds like a stretch		Will delete piece-wise	✔
p.2: "ROC" stands for "Receiver Operating Characteristic" p.2: "gaussian" should be "Gaussian" (throughout) p.2: "euclidean" should be "Euclidean" (throughout)		Will do	✔
p.3: While one could argue that "divisive" strictly can relate to the mathematical process of division, it is way more typically used in relation to disagreement or hostility between people. I find it's use in this context distracting		Will change to **normalization by division**	✔
p.2 (literature survey): are all the performance figures quoted in the survey performed on the same dataset using the same protocol? If not, then they are misleading	The studies highlighted in the literature review are chosen due to their similarity to this one based on methods used. Most used the DRIVE data set and a few used the STARE data set.	The study has been extended to include the STARE data set.	See Authors comments.
p.3 (equation 1): the independent variables on the LHS and the RHS are different, so the equation makes no sense		Will change to x and y on LHS to x' and y' respectively.	✔
p.3 (Gabor filters): it is stated that the Gabor filter response is a complex number, but the equations given are real valued. Which is used in the subsequent portions of the work?	Complex numbers consist of a real and imaginary part. In this study (as is the case in other studies) only the real part is used. Equation is 1 is described in the paper as the real component of the gabor function.	Will explicitly state that the real component is used in subsequent portions of the work.	✔
p.3 (equations): Equations should be formatted and punctuated as part of the text. See Mermin, "What's wrong with these equations", Physics Today, 1989 (all other equations too)		Will add punctuation marks where appropriate.	✔
p.3: "are prone to having a gaussian distribution". The word		Will change to “usually have a	✔

"prone" implies a regrettable vulnerability and is inappropriate in the context used		gaussian distribution”	
p.4: "Illumination variance" - the usage of "variance" in the field usually refers to the formal statistical variance of a distribution or of a set of numbers. "Variation" would perhaps be less objectionable		Will change to variation	✔
p.4: "K_{w \times w}" - how is w chosen? Presumably its value changes with wavelength?	It is set to 21 pixels to allow for detection of the largest vessels.	Will state this.	✔
p.4: "are strictly as a result of linear detection" should be "are strictly a result of linear detection". However, I also don't know what "linear detection" means in this context		Will change. Will use line detection instead.	✔
p.4: Algorithm 1 and Algorithm 2, while they do specify the process, are extremely simple and it is really not necessary to present the methods in such a laborious form.	We believe this enables us to present the method in a replicable manner. We prefer to keep it this way.		See Authors comments.
p.4: "sum of positive and negative responses within a window sum up" - "within a window" is imprecise and possibly confusing		Will remove within a window .	✔
p.4: the usual way of making a kernel that leads to offset invariance is to subtract the mean. It is not clear to me what the methods presented in Algorithm 1 and 2 are trying to do in this context. To normalise the scale it is also typical to make the kernel have an L2 norm (sum/integral of squared values) of unity. Making the "area under the gaussian curve of the gabor kernel sum up to one" doesn't quite achieve this either. In short, I'm not sure what Section 2.3.2 is trying to achieve, and I'm not really happy with what is being done there	L2 norm simply scales all the values to [0,1]. This does not solve the problem of illumination variation. The variation will still be there but only scaled down to a smaller range. More over the negative responses of the gabor filter will be lost. The objective is to make areas of constant illumination have a response of zero. In any case L2 normalization was explored and it gave very poor results.	Will elaborate on the purpose of the algorithms.	✔
p.4: "when it's pixels have a high correlation" - with what?	As in correlation co-efficient / highly co-related.	Will change to have high similarity with each other.	✔
p.4: "The algorithm accepts an input image that has been normalised using equation 7" - but equation 7 has nothing to do with the input image		Will change to: an input image that has been convolved with the scaled gabor equation in equation 7.	✔
p.4: "This normalisation ensures that the area under the gaussian curve of the gabor kernel window sums up to one". Really? Either provide reference or prove		Will elaborate and provide reference.	✔
p.5: "The median is used as a base". Why? What criterion?	When the median is computed, the value returned is	Will elaborate.	✔

	one of the actual values within in a window. The resulting filtered image is therefore likely to have high resemblance of the original image. Other based could be used as well.		
p.5 (Algorithm 3): The presentation is inconsistent. "for i:=-w/2 to w/2" only makes sense if w is even, while it is clearly envisaged that the kernel will have an odd dimension.	w is always odd in the code. Because of Integer division (high level programming languages discard the fraction when performing integer division), the presentation only makes sense if w is odd.		See Authors comments.
p.5 (Algorithm 3): The word "offset" in the algorithm is formatted as variables in a math environment. Find out about "\text{}" and use it here and elsewhere as appropriate	It is a variable.		See Authors comments.
p.5: Equation 8 uses p_t while equation 9 uses p_i for the same quantity. Why the inconsistency?	Typo	Will change p_t to p_i	✔
p.5 (footnote 1): "i" should be formatted as math to be consistent		Will do	✔
p.5: "The advantage of VET...is that it is capable of thresholding both unimodal and bimodal histograms". This statement makes no sense. A flip of a coin is able to do the same. By what criterion, and using what assumptions and principles can this be done?	Although this advantage is clarified in the cited text on VET we have no problem explaining it further.	Will elaborate on VET and highlight effective thresholding in the context of fore and background segmentation.	✔
p.5: "Automated segmentation is an unsupervised classification problem and hence its different possible outcomes can be illustrated using a contingency table". I don't understand the logic here - why is it significant that the problem be unsupervised for this to be true?		Will delete unsupervised	✔
p.6: "prone" is wrong in this context		Will change to: its boundary is likely to be detected as an vessel edge.	✔
p.6: "Blood vessels have stabler inter-pixel gradients" - than what?	Than non-vascular tissue.	Will change	✔
p.6: "region.The" should be "region. The"		Will change	✔
p.6 (figure 3): the similarity in the results for (c) and (d) makes me question whether it's worth reporting on procedures that are obviously almost identical	The vessels in d are thicker. This makes a difference depending on whether an image has predominantly thick or thin vessels.		See Authors comments.
p.6: In all these results there is no indication of how the (scalar) Gabor filter response used in the thresholding is obtained from		Will elaborate	✔

what is presumably the output of a set of Gabor filters.			
p.7: Figure 4 is horrible: while it contains information for three different wavelengths the plot is formatted in a way that hides this. Frankly, a table would be better. Also, what about orientation?	We do not concur. Non the less we will go with the suggestion.	Will change to table	✔
p.7: Figure 5 would similarly probably be better represented by a table - it is not contributing clarity	We do not concur. Non the less we will go with the suggestion.	Will change to table	✔
p.7: "When each pixel is filtered, all wavelengths within the range are considered". What range? Discretisation? And what about angles?		Will elaborate	✔
p.7: "The ROCs in Figure 6 show the trade-off between sensitivity/true positive rate and FPR for each normalisation method when the image...is manually thresholded". What? For a manual threshold there is only one value of TPR and FPR. Presumably this curve is generated by varying the threshold for each of the normalisations	The threshold for each normalization is varied.		See Authors comments.
p.7: "This shows that they are all viable for segmentation". What is that supposed to mean?	They are all viable for use in retinal image segmentation.	Will change accordingly	✔
p.7: "the other methods have angular curves". Something here sounds odd to me		Will change curves to graphs	✔
p.7: The results in Section 3.2 are all given without any mention of the dataset, in particular the number of images and how the ground truth is obtained. While these details are mostly given in Section 4 that makes no sense in the context of the presentation		Will include a description before section 3.2.	✔
p.8: "The images are divided equally into training and testing samples". Since the method presented here doesn't use any training, presumably all images were used for testing?	Only the test set was used for testing. The reason for this is explained in the paper.		See Authors comments.

Reviewer B

Reviewer's Comment	Author's Comment	Proposed Action	Done
The paper focuses on the improvement of the segmentation of retinal blood vessels by reducing background illumination variance. For this normalization step Gabor filtering is proposed considering several possible approaches, like base-offset encoding and zero-integral kernel techniques.	Correct	N/A	N/A

-- "gabor" should be started with capital letter: Gabor		Will do	✔
-- "fundas image" should be: "fundus image"		Will do	✔
-- There are some typos/linguistic bugs, thus, a careful revision is needed. E.g.: "Although gabor filters are an effective line detection tool, they have the draw back..."		Will change to they are dependent on	✔
-- It Would be nice to see the actual differences in the implementation of Algorithms 1 and 2 with respect to the ones given in [20] for segmented images.	Algorithm 2 is a re-inactment of [20] while algorithm 1 is a modification thereof.	Will ellaborate on this.	✔

Reviewer C

Reviewer's Comment	Author's Comment	Proposed Action	Done
The authors propose a method for vessels segmentation in retinal fundus images based on normalization of Gabor filter responses and Valley Emphasis Thresholding (VET). **The novelty of the work is in the proposal of new normalization methods.**	Correct	N/A	N/A
- In Equation (1), the left side contains x and y while in the right side there are x^\\prime and y^\\prime. It should be fixed.		Will change to x and y on LHS to x' and y' respectively.	✔
- Around the paper it is used the word 'fundas' instead of 'fundus'.		Will change	✔
In the Literature Survey section, results from the other methods are reported but it is not specified on which data set. The data set is only specified next in the text. In general, it is preferred to report the results of other methods in the discussion section.	We prefer to retain the results from other methods in the survey as it makes for a comprehensive review of what has been done and how well.	Will specify data sets used by other methods.	✔
In the Zero-Integral normalization subsection, it would be appreciable to more strongly highlight the differences between the two algorithms and when one is preferable on the other one. The pros and cons of the two algorithms should be discussed.		Will discuss.	✔
In the Evaluation subsection, it is stated that " Automated segmentation is an unsupervised classification problem and hence its different possible outcomes can be illustrated using a contingency table 1.". I do not agree with this sentence, since automated segmentation can be achieved also using supervised classification. Segmentation is a binary classification problem and this is the reason why for each pixel four possible outcomes are possible.		Will replace unsupervised with binary.	✔
Moreover, it is not necessary to specify as an evaluation metric the FPR, since it is computed from the Specificity as "1-Sp".		Prefer to change equation of FPR to 1-Sp	✔

Then, it is stated that the ROC curve is achieved by varying the regulated parameter. I can imagine that this parameter is the threshold value, but it should be clarified.		Will do	✔
In the Automatic thresholding section, it is not clear to me the meaning of Table 2. The sentence " Table 2 shows that the VET method predicted the most optimal threshold for both the median-offet and mean-offset methods." is not explanatory of what the Table 2 reports.		Will elaborate on table 2	✔
The experiments have been carried out on one publicly available data set widely used for benchmarking of segmentation algorithms, namely the DRIVE data set. In literature there are other standard data sets that should be used to provide a deeper validation of the method: the STARE and the CHASE_DB1 data sets. The authors should, at least, provide experimental results on the STARE data set, which is older than the DRIVE data set and contains an higher number of images of the retina that show signs of pathologies. The CHASE_DB1 contains, instead, higher resolution images with some challenges due to the light reflex in the middle of the vessels. It could be interesting to investigate the response of the proposed normalization algorithms on such images.		Will extend experiments to include STARE data set.	✔
The results comparison is incomplete. There are many other methods, also cited in the state of the art section, that achieve much higher results that the method considered for the comparison. For instance the methods of Mendonca and Campilho (2006), Ricci and Perfetti (2007), Al-Rawi et al. (2007) or other methods for which I report the references should be included in the analysis:	The cited methods have been chosen because they use the DRIVE database. Due to the number of previous studies, an exhaustive list is impractical. An incomplete list will inevitably leave out some studies. The chosen methods are merely to show that our results are comparable and not necessarily superior. Generally, supervised methods will have higher results than unsupervised. Most of the studies cited by the reviewer are supervised.	The reader will be referred to [9] for a detailed list of previous results.	✔