Depth discontinuity-based cup segmentation from multi-view colour retinal images

Gopal Datt Joshi, *Member, IEEE,* ,Jayanthi Sivaswamy *Member, IEEE,*, S. R. Krishnadas .

***Abstract*—Accurate segmentation of the cup region from retinal images is needed to derive relevant measurements for glaucoma assessment. A novel, depth discontinuity (in the retinal surface) based approach to estimate the cup boundary is proposed in this paper. The proposed approach shifts focus from the cup region used by existing approaches to cup boundary. The given set of images, acquired sequentially, are related via a relative motion model and the depth discontinuity at the cup boundary is determined from cues such as motion boundary and partial occlusion. The information encoded by these cues is used to approximate the cup boundary with a set of best fit circles. The final boundary is found by considering points on these circles at different sectors using a confidence measure. Four different kinds of datasets ranging from synthetic to real image pairs, covering different multi-view scenarios, have been used to evaluate the proposed method. The proposed method was found to yield an error reduction of 16% for cup-to-disk vertical diameter ratio (CDR) and 13% for cup-to-disk area ratio (CAR) estimation, over an existing monocular image based cup segmentation method. The error reduction increased to 33% in CDR and 18% CAR with the addition of a third view (image) which indicates the potential of the proposed approach.**

***Index Terms*—Cup segmentation, Multi-View Retinal Images, Glaucoma, Cup-to-Disk Ratio, Optic Disk.**

I. Introduction

Screening for retinal diseases is crucial for early detection of disease, and subsequent treatment to avoid preventable vision loss [1]. Glaucoma is one of the common causes of irreversible vision loss and is characterised by the progressive degeneration of optic nerve fibers. The amount of nerve fibers exiting the eye from different parts of the retina is represented by an annular region between the optic disk (OD) and cup boundary as depicted in Fig. 1(a). This region is known as neuro-retinal rim. The loss in optic nerve fibers leads to an enlargement of the cup region or thinning of this neuro-retinal rim called *cupping* which is an important indicator of glaucoma progression. Various indicative parameters such as cup-to-disk vertical diameter ratio, cup-to-disk area ratio, etc. are measured to quantify the amount of cupping.

Manuscript received October 10, 2011; revised December 31, 2011; accepted January 30, 2012. This work was supported by the Department of Science and Technology, Government of India under Grant SR/S3/EECE/0024/2009.

Gopal Datt Joshi and Jayanthi Sivaswamy are with the Centre for Visual Information Technology, IIIT Hyderabad, Hyderabad 500032 India (e-mail: gopal@research.iiit.ac.in, jsivaswamy@iiit.ac.in).

S. R. Krishnadas is with Aravind Eye Hospitals, 625020 Madurai, India (e-mail: krishnadas@aravind.org).

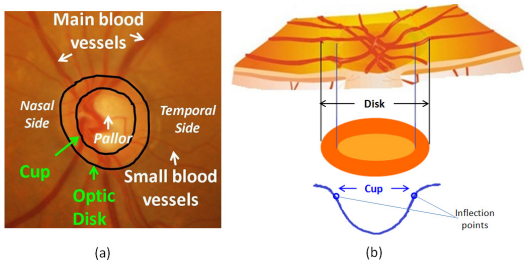

Fig. 1. A monocular retinal image and a 3D visualisation of the surface. (a) rendered retinal surface and (b) the corresponding 2D projection and cross-sectional depth profile.

A colour retinal (fundus) image (CFI) is a 2D projection of the retinal structures and the OD appears in the CFI as a bright circular or elliptical region partially occluded by blood vessels (Fig. 1(a)). From a computational point of view, accurate segmentation of OD and the cup region from CFI are the two fundamental tasks which are required to quantify the amount of cupping, to estimate the indicative parameters of glaucoma. OD segmentation methods [2] [3] have gained reasonable amount of maturity, and are capable of providing reliable OD boundaries. Whereas, the detection of cup boundary from a CFI is still a challenging task as depth is the *best* marker for the cup (Fig. 1(b)) which is lost in the 2D projection. Earlier work on cup segmentation can be broadly divided into two categories based on their choice of CFI modality: monocular or stereo. In monocular case, various 2D visual features such as kinks (vessel-bend on boundaries) [2] [4] [5], discontinuity in the pallor region [6] [7] and shape priors [7] have been exploited to estimate cup boundary. However, reliable detection of these features is challenging mainly due to the high structural and photometric uncertainties around the cup region. In contrast, stereo imaging helps to get a richer representation of the objects in a scene and hence has been used to derive depth information for segmenting the cup region.

Two different cup segmentation strategies have been explored (to the best of our knowledge) using stereo image pairs. In the first strategy, 2D visual features are integrated from stereo pairs [8] [9]. Here, features are computed using simple and complex cell models, and the cup region is obtained using a pixel based classification. The second strategy computes depth/disparity from stereo image pairs via stereo correspondence [3] [10] [11] [12] [13] [14] [15]. Different correspondence techniques (pixel-based or region-based) have been proposed for the task of cup segmentation. Camera calibration [13] [11], image pair registration [10], etc. have been employed to improve the robustness of the correspondences.

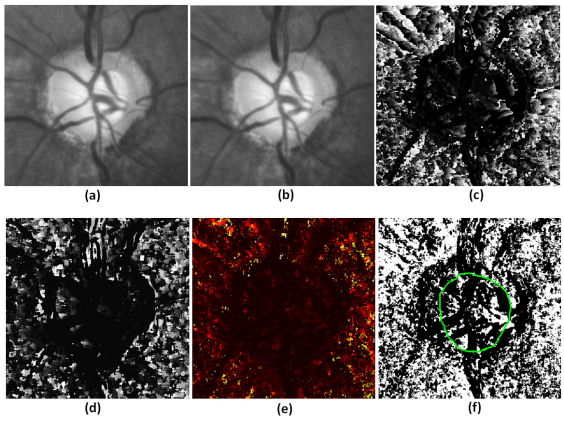

Fig. 2. Stereo correspondence results on an image pair obtained via an ad hoc procedure. a) Left image; b) right image; The extracted disparity maps using c) feature-based, d) region-based correspondence techniques; e) the disparity map obtained after integration; f) Binary map of first set of matched points (white) with the expert marked cup boundary shown in green colour.

The computed depth (or disparity between stereo pairs) map is further refined by using imaging configuration details such as baseline distance, camera viewing angle and position, in the stereo pair. The desired cup region is then obtained by thresholding the computed depth/disparity map.

The scenario in developing countries is dominated by the use of monocular fundus cameras. This is mainly due to the high cost differential between fundus cameras with monocular versus stereo imaging capability. Sequential acquisition of images by manually varying the position of the monocular camera position is an ad-hoc procedure adopted to acquire stereo image pair. Such image pairs are displayed on computer monitor and a hand-held instrument called stereoscope is used for optically fusing them and create a 3-D (depth) image examination experience. While such image pairs may be adequate to get 3D perception with manually adjusted settings, they are inadequate for automated extraction of depth information. This is because the manual shifting of the camera in this procedure results in *inconsistencies* in the degree and direction of stereo-separation, the plane of focus and illumination. Therefore, earlier reported methods on retinal stereo correspondence cannot be applied directly on such image pairs. This is illustrated by applying the stereo correspondence technique in [14] on a sample pair of images which are shown in Fig. 2 (a) and (b).

A computational step similar to that presented in [14] was employed prior to the correspondence step. Two disparity maps were first obtained using feature-based (intensity and gradient) and region-based (normalised cross correlation) information. In order to determine a reliable disparity map, the disparity maps obtained from feature and region based matching techniques are integrated. This is done by first retaining only those points which have similar disparity value assigned by both the techniques and then including other points in multiple iterations, using a proximity constraint on the *first* set of points. The disparity maps obtained by two different matching techniques are shown in Fig. 2 (c) and (d), respectively. Fig. 2(e) and (f) show the disparity map after integration and the binary map of the first set of reliably correspondences (in white), respectively. The expert marking for the cup region is overlaid in green colour to aid assessment of the result. It can be seen that only a sparse set of reliable correspondences is obtained, particularly inside the cup region. This provides a weak support for extracting the cup region. Moreover, the cup region is characterised by low disparity values (as indicated by the abundance of dark pixels in Fig. 2(e)) due to the shallowness of the cup region. The sparseness on the obtained disparity map indicates the difficulty in obtaining reliable disparity values for *all* points which belong to the cup region. Therefore, dealing with image pairs acquired via an ad-hoc procedure needs a fresh look and finding an alternative to stereo-correspondence based strategy for cup segmentation is therefore of interest.

In our work, we consider such image pairs as two views of the retina acquired at two different time instances. The multi-view perspective treats the image pair as source of information of a 3D structure but not solely on the basis of disparity. Thus, it is possible to relax the strict-stereo pair requirement. However, the multi-view perspective increases the complexity and challenges for the task of cup segmentation. In this paper, we consider the cup segmentation problem with multi-view images as input and shift the focus from the cup region to cup boundary in order to address the additional challenges and requirements posed by this multi-view perspective. This focus shift allows reformulating the cup segmentation problem from one of estimating the depth value at each point in the OD region, to that of localising discontinuities in depth in the same region. Moreover, the inclusion of multiple views (more than two images) can be used as an additional source of information to further improve the accuracy of cup segmentation. In the next section, a detailed explanation of the proposed formulation is provided.

II. Background

To the best of our knowledge, a boundary (rather than region) based strategy has not been explored to address the cup segmentation problem. There is a good body of work related to 3D scene geometry in computer vision literature [16] [17] [18] which can be utilised to develop such a strategy. The first task at hand is to define ways in which depth discontinuity, which characterises cup boundary points, can be computed from given multi-view image data. We begin by noting that there is a relative motion between the camera and the eye. This relative motion arises either because a) the camera moves in front of a stationary eye to capture the two images (views) or b) the eye involuntarily moves during the acquisition of these images simultaneously. We attempt to extract depth discontinuities by exploiting information available in the given views due to this relative motion. These scenarios describe both strict stereo and the ad hoc procedure of sequentially acquiring the images. In both scenarios, we assume that the motion parameters are unknown.

Fig. 3 illustrates the change in a projected image point corresponding to a relative motion between the camera and a scene (retinal surface). If a point P on a surface moves with a velocity v_o, then the imaged point p can be assigned a vector v_i to indicate its movement on the image plane. The changes in

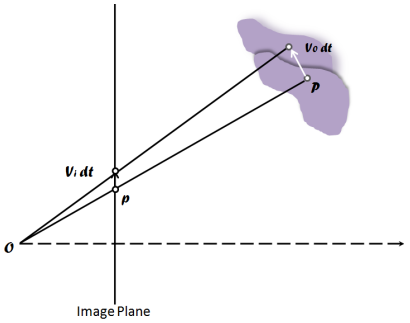

Fig. 3. Displacement of an image point due to relative motion between camera and an object.

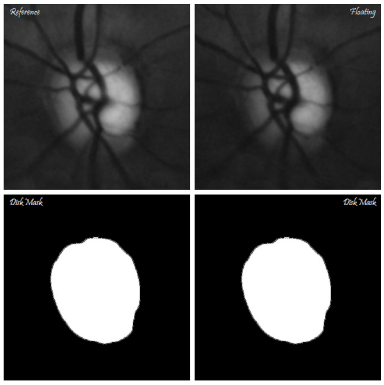

Fig. 4. Sample pair of retinal images or 2 views (green channel) (top row) and OD masks (bottom row) derived using [2].

the image points encode information about the camera motion and the surface geometry. The precise estimation of motion parameters and 3D scene geometry is a challenging problem attracting considerable attention in computer vision. However, in the problem at hand, we are only interested in characterising and subsequently detecting depth discontinuities associated with the cup boundary and do not require complete motion parameter estimation or 3D reconstruction.

We propose to localise the depth discontinuities using multiple cues, drawn from the image motion field, which are listed below.

1) *Discontinuity in motion field:* Inconsistencies or abrupt changes in image motion are primarily associated with object and depth boundaries and also indicative of partial occlusions. [19] [20]
2) *Partial occlusions:* Some points are visible in only one of the two images due to partial occlusion. This depends on the camera configuration and on the depth variation in the scene. Partially occluded points are directly associated with depth discontinuities [20] [21].
3) *Contiguity of discontinuities:* Depth discontinuities are spatially contiguous as they arise from objects and their boundaries [19]. When an underlying object shape is known, this property can be used to estimate object boundaries at locations even in the absence of motion evidence to handle errors in localisation of depth discontinuities.

Information on these cues can be obtained from the image motion. A solution based on this proposal for cup segmentation from two different views of retinal surface is presented next.

III. Proposed Solution

We begin with an overview of the proposed solution. The algorithm for motion (displacement) field estimation needs to handle the arbitrary relative motion between the camera and the underlying non-rigid retinal structures. We choose the Demon's registration algorithm [22] to find the displacement field between two images. Here, image registration is formulated as an optimization problem that aims at finding the displacement of each pixel in order to get a reasonable registration between two images. Abrupt changes in the resulting displacement field are used to get potential motion boundary locations. The partial occlusion locations are extracted from an error image which is derived by identifying point-wise mismatch between a reference and the registered floating image. The potential depth discontinuity points associated with the cup boundary are obtained next by eliminating discontinuities that arise due to vessels.

A. Pre-processing

The given pair of retinal images are divided into a reference image R and a floating image F. The location of OD, the region of interest, is obtained from both R and F using a circular Hough transformation-based approach presented in [2]. Next, an image region of sized 800×800 is extracted centered on the OD, for each image. We consider the green color plane for further processing since it has a good image contrast. A mask for the OD is determined by employing a modified Chan-Vese model [2]. A sample pair of reference and floating images and the computed OD masks are shown in Fig. 4.

B. Displacement field computation

Let r and f denote the pixel intensity from reference image R and the intensity from floating image F respectively, at a given location *p*: (x, y). The displacement (velocity) u needed for pixel *p* in R to match the corresponding pixel in F is defined as [23] [24]

$$u = \frac{(f-r) \bigtriangledown r}{|\bigtriangledown r|^2 + \alpha^2 (f-r)^2} + \frac{(f-r) \bigtriangledown f}{|\bigtriangledown f|^2 + \alpha^2 (f-r)^2} \qquad (1)$$

where $u = u_o.e^{j\theta}$, and $\bigtriangledown r$ and $\bigtriangledown f$ are the gradients of the reference and floating image, respectively. There are two types of forces: an internal edge-based force $\bigtriangledown r$ or $\bigtriangledown f$ and the external force $(f - r)$. The weight α helps in adjusting the force strength. An intensity-based registration energy is defined between reference and floating image to get u for each image pixel. This energy term consists of the squared pixel distance as a similarity function and the squared gradient of the transformation field as smoothness regularization. The optimization is performed over the entire space of displacement fields to optimally align the two images. The resulting displacement field is later used to derive desired motion cues.

The registration procedure is summarised as follows:

- Start with a spatial transformation S describing the translation in x and y direction of every pixel from its original position
- Iterate until convergence

1) Given S, compute a corresponding update field U by minimising

$$E(U) = ||R - F \circ (S + U)||^2 + \frac{\sigma_i^2}{\sigma_x^2}||U||^2 \quad (2)$$

where U is the update of S, $\circ$ image transformation. σ_i and σ_x account for the noise in the image intensity and for a spatial uncertainty in the correspondences, respectively.
2) Let $U \leftarrow K_{fluid} * U$. The convolution kernel K_{fluid} is a Gaussian kernel used for regularisation.
3) Let $S \leftarrow S + U$
4) A diffusion-like regularisation is included, with $S \leftarrow K_{diff} * S$ where K_{diff} is a Gaussian kernel.

Now, energy E for the pixel p can be rewritten in terms of the pixel intensities from the reference image R and transformed image $(F \circ S)$, u the x and y direction translation updates of the pixel from U, and with $\bigtriangledown f$ the image gradient at pixel p.

$$E(u) = ||r - f + u \bigtriangledown f||^2 + \frac{\sigma_i^2}{\sigma_x^2}||u||^2 \quad (3)$$

Accordingly, the error gradient,

$$\bigtriangledown E(u) = 2(\bigtriangledown f)^T(r - f + u \bigtriangledown f) + 2.\frac{\sigma_i^2}{\sigma_x^2}u \quad (4)$$

Assuming that E is minimum at $\bigtriangledown E(u) = 0$, we can get required update:

$$u = \frac{r - f}{|| \bigtriangledown f||^2 + \frac{\sigma_i^2}{\sigma_x^2}} \bigtriangledown f \quad (5)$$

For a detailed explanation of the optimization procedure, refer to the original work [25].

The computed displacement field at every point is used to derive a registered image (denoted as F_R) from F. Two types of information are derived from $u(x, y)$ and F_R: i) local motion inconsistency which signals boundary points. This is derived from the displacement information, specifically, u_x and u_y, the projection of vector u in x and y directions and magnitude u_o defined as $u_o = \sqrt{u_x^2 + u_y^2}$; ii) registration error (e_m) information. This is computed by taking the absolute difference between R and F_R. Locations with high registration error signal partial occlusion points. This can be reasoned in the following way. The registration error obtained at a point can be due to two factors: a) photometric changes between the two views/acquisitions, and b) partial occlusions. As mentioned earlier, partial occlusion usually occurs at points of depth discontinuities and in the absence of true matches, mismatches occur on either side of the depth discontinuities. Since obtained matches for such region pixels are always false, it leads to relatively high valued regions in the registration error map. Fig. 5 shows the four maps u_x, u_y, u_o and e_m derived for the sample image pair given in Fig. 4. Only absolute values

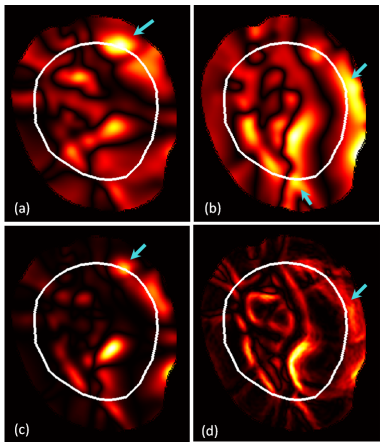

Fig. 5. Evidence maps for the image pair in Fig. 4. a) u_x, b) u_y, c) u_o, and d) e_m images. The cup boundary (in white) provided by an expert is overlaid for comparison.

are considered. These are visualised as heat maps with yellow and black representing high and no responses, respectively. The ground truth b_g, which is the expert marked boundary, is overlaid to help analyse the obtained responses.

At a first glance, the e_m map looks very cluttered due to depth discontinuities introduced by the vessels inside the cup region. The occlusion points can be seen at the edges of vessels which arise due to a relative edge shift between two views. However, its integration with the other maps can help to get desired partial occlusion associated with the cup boundary. A close look reveals that b_g and locations of high response in u_x, u_y and u_o are often aligned with expert marked cup boundary, highlighted by blue arrows in Fig. 5. The e_m values are high outside the b_g (bottom right) indicating presence of occlusion points. The co-occurrence of b_g and the obtained high responses is apparent at different locations. In theory, it can occur on either side of b_g depending on the direction of camera motion. For instance, in Fig. 5 there are sufficient amount of evidences in the right side as compared to the left side which implies that camera was moved to the right relative to the first camera position to acquire the second image. Thus we conclude that these maps derived from two views provide evidences for the location of a set of points along the cup boundary. Based on the strength of evidence at a point, a set of best fit circles centered at the middle of the optic disk can be derived as estimates for the cup boundary.

C. Approximating the cup boundary

The approach taken for finding a potential cup boundary point is based on evidence gathering from all maps. A high response at a p_i in *all* the 4 maps is deemed to be a strong evidence for p_i to be a point on depth discontinuity. Towards this goal, the 4 maps are first normalised to a range [0,1] and then thresholded (value=0.35) to derive binary maps of high response points. These correspond to true candidates as well as false ones in the interior which arise due to the vessels within the cup region. The maps are therefore pruned to reject all these interior points with the help of a rough, cup region-mask derived using the pallor information (see [6] for details). Since pallor is a bright region in a retinal image and always

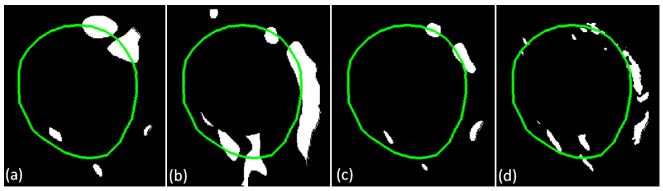

Fig. 6. Pruned binary evidence map obtained for a) u_x, b) u_y, c) u_o, and d) e_m images. The retained evidence points are in white and the expert-marked cup boundary is in green.

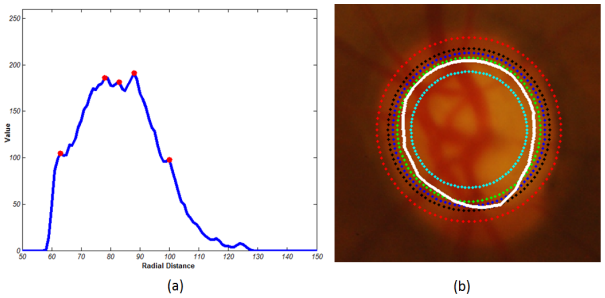

Fig. 7. a) Combined histogram H_c and prominent peaks at r_i, b) Cup boundary (white color) with circular contours of radii r_{ci}. The top five peaks from H_c were used to get these circular contours.

inside the cup boundary, this pruning only eliminates a subset of outliers. Fig. 6 shows the pruned binary evidence maps.

Following this pruning, we gather evidence that a point qualifies to be a point on cup boundary. This is done by computing a distance histogram H_c as shown in Fig. 7(a) which is obtained by counting the number of co-located points across the 4 maps. The radial distance from the center of OD is taken to indicate the location. The local maxima in H_c above certain minimum value represent the locations with maximum evidence or agreement between motion discontinuity and occlusion information. Local maxima above an empirically determined threshold of 50 are considered. These peaks are highlighted as red dots in Fig. 7(a). The locations r_i corresponding to these peaks are then used to define multiple circles of radii r_i which are best fit circles for different cup boundary points. Fig. 7(b) shows the obtained circles for the first 5 peaks, along with b_g. It is seen that there is a good alignment between a circle and the b_g in each angular

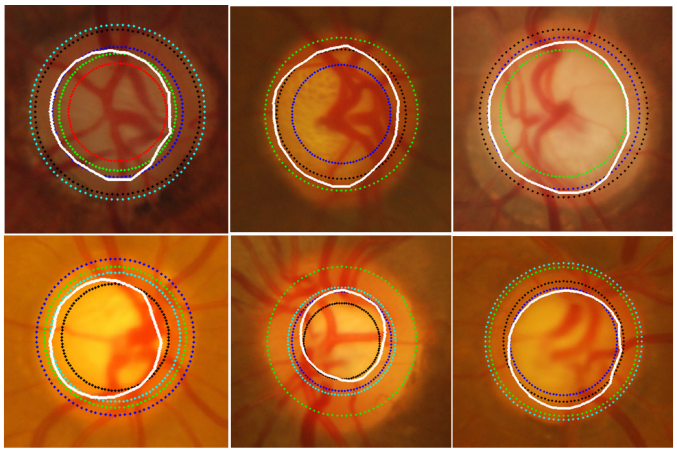

Fig. 8. Six examples illustrating cup boundary with estimated circular contours.

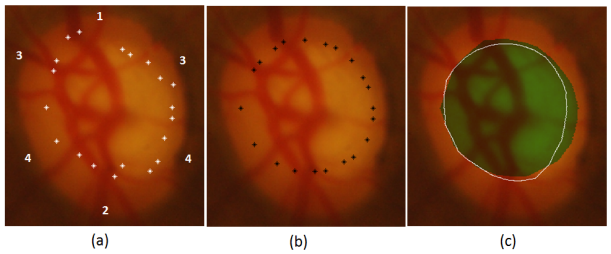

Fig. 9. Cup boundary extraction example. a) Set of points selected in different image sectors, b) result of smoothing the previous set of points, c) extracted cup (shaded in green colour) along with expert marked boundary (shown in white).

sector. This validates our assertion that these four maps are capable of encoding depth discontinuities associated with the cup boundary. Fig. 8 shows similar circular contours obtained for six more image pairs (only one of each pair is shown) with a variety of OD and cup appearances.

D. Estimating the final cup boundary

The difficult task that remains is that of estimating the final cup boundary b_c as a set of points $p : (x, y)$ selected from the multiple circular contours (radii) obtained above. In other words, the task of identifying right circular contour from the derived contours which best represents cup boundary in a particular sector. We do this by computing a confidence value c_p for each location and choosing points sector-wise with i) high confidence value and ii) backed up by evidence from the e_m map. The confidence value c_p for a point (x, y) is defined as

$$c_p(x, y) = w_1 * u_x(x, y) + w_2 * u_y(x, y) + w_3 * u_o(x, y); \quad (6)$$

where $w_1 = 0.2$, $w_2 = 0.2$, and $w_3 = 0.6$ are the weights that were *empirically* fixed. A higher weight is assigned for the u_o because it captures the displacement magnitude of the point. Next, for each sector, the point with maximum value of c_p is selected and retained if, in its vicinity of 10 pixels, there is a high response in e_m otherwise next maximal point is selected. In our work, 36 angular sectors were considered. Fig. 9(a) shows the obtained points in different sectors. According to the contiguity of depth discontinuities constraint, the cup region should give rise to continuous depth edge along its boundary. Hence, a local radial smoothing is done to obtain the final b_c and to handle outliers. A procedure shown in Algorithm 1 summarises the final cup boundary derivation procedure.

Fig. 9 shows the outcome of these steps on a sample image along with the final cup region. The results indicate that the algorithm is able to capture the boundary points successfully except for the lower sector (see Fig. 9(c)). In this sector, the local deformation in b_g is captured well in u_y. The selection stage picks the high response arising from vessels inside the cup boundary instead of true one. Deriving a common set of selection criteria to localise points of true depth continuity from these two dimensional evidence maps is challenging. The nature of difficulty in accurate localisation is also evident from the inter-observer variations seen in expert marked cup boundaries. Fig. 10 shows four sample segmentation results against 3 experts. It is seen that at different locations, a

Algorithm 1 Final cup boundary derivation procedure

Get all points in the vicinity of multiple circles defined by r_i
For each point, compute c_p
Scan each sector in steps of 10°
In each step, select point with the maximum value of c_p
if e_m response is high in local vicinity **then**
    Include point in final points
else
    Look for next maximal point and repeat step 5
end if
Perform radial smoothing on the final boundary points
Get the segmentation for the cup region

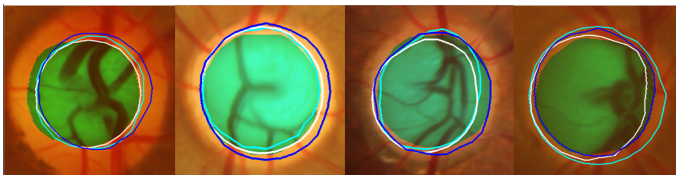

Fig. 10. Sample cup segmentation results along with ground truth obtained from 3 experts. Green shaded region denotes the cup region obtained by the proposed method; expert markings are shown using coloured contours.

reasonable amount of disagreement on the cup boundary is present among experts. However, the selection procedure used in the proposed method can become reasonably accurate if images from multiple (more than two) views are used to gather the evidences.

IV. Experimental Results

The main objective of our experimentation is to assess the ability of the proposed formulation in extracting cup boundaries from 2 views/images obtained under different scenarios. Four different kinds of dataset represent these scenarios. Two variants of the proposed solution were analysed: 1) Variant-1 (v-1): all processing was carried out on the entire OD region, and 2) Variant-2 (v-2): processing was restricted to vessel points. These were found by masking the four evidence maps with the vessel mask obtained by first segmenting vessels within OD. The segmentation step uses the method in Garg et al. [26]. This variant is meant to capture 2D visual cues (kinks [4]/relevant vessel bends [2]) typically used by the experts while marking cup boundary.

A. Data Specification

1) Dataset-1: Image pairs were generated using a ray-traced model of the camera-eye-retina system presented in [27] by laterally translating the camera to the left *and* to the right by 5 and 10°. This angle is calculated between the centers of the objective lens and the cornea. In this model, a 3D retina is simulated with a 2D texture of a real fundus image placed at back of a virtual eye, and image pairs are captured by translating the virtual camera in a horizontal direction by different amounts. Thus, this simulates imaging with variable camera translations. All images are obtained at a resolution of 1024×768. This can be considered as an example of controlled acquisition of multi-view retinal images where imaging variations are minimal such as patient eye movement, inter-image illumination variations, etc. Three samples of the generated image pairs are shown in Fig. 11.

2) Dataset-2: This dataset was gathered from 15 Normal and 15 Glaucomatous eyes using monocular fundus camera. Each collected data tuple consisted of an image pair, an associated OCT data and boundary (OD and cup) markings (on CFI) from one glaucoma expert. No imaging constraint was imposed during image acquisition except for a fixed field of view (30°) and image resolution (2896×1944 pixels.). The camera motion, in terms of the degree and direction of displacement, varied across the acquired image pairs and are unknown. The OCT and expert markings serve as two different sources of ground truth in the assessment.

3) Dataset-3: This dataset was gathered from 33 normal and 105 glaucomatous (total of 138) eyes using monocular fundus camera. The classification between normal and glaucomatous eye was performed by a glaucoma faculty by reading retinal image. A case was declared normal if it was not glaucomatous. For each image pair, markings were collected from three glaucoma experts, referred to as Expert-1, Expert-2 and Expert-3 with experience of 3, 5 and 20 years, respectively. The imaging specifications were same as in Dataset-2. An average ground truth (Expert-4) was derived for each image by averaging boundaries obtained from three experts [2]. This dataset (one image from an image pair) has been used for the evaluation of OD and cup segmentation from monocular retinal images [2].

4) Dataset-4: This dataset had image pairs acquired with a fixed degree and direction of camera motion. These image pairs are captured by monocular Heidelberg Retina Angiograph (HRA) equipment. Red-free (grayscale) image pairs are captured by a fixed camera with translation which was produced automatically by the equipment's firmware. The captured image resolution is 768×822 with 30° field of view. Three sample image pairs are shown in Fig. 13.

B. Results

1) Dataset-1: The effect of variable degree of camera motion was studied on the v-1 algorithm. Fig. 11 shows the cup segmentation results obtained on this dataset. The consistent results across image pairs show the ability of the proposed method to deal with camera translation range from -10° to 10°.

Fig. 11 shows 3 sample image pairs obtained synthetically. A lateral shift can be seen between the two views (top and middle row). In general, a change in the view angle can cause a change in the appearance of the OD and cup regions. This can be verified to some extent by the computed OD and cup boundaries. The OD boundary extraction is using the method in [2] which does not use depth information. The extracted cup boundary is consistent which indicates the robustness of the depth discontinuity based formulation.

2) Dataset-2: Two parameters, namely, the cup-to-disk vertical diameter ratio (CDR), and cup-to-disk area ratio (CAR)

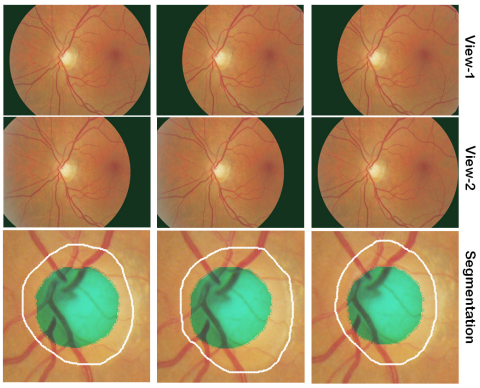

Fig. 11. Sample synthetically generated image pairs (first two rows) and cup segmentation results in green colour(last row). The OD boundary is shown in white colour.

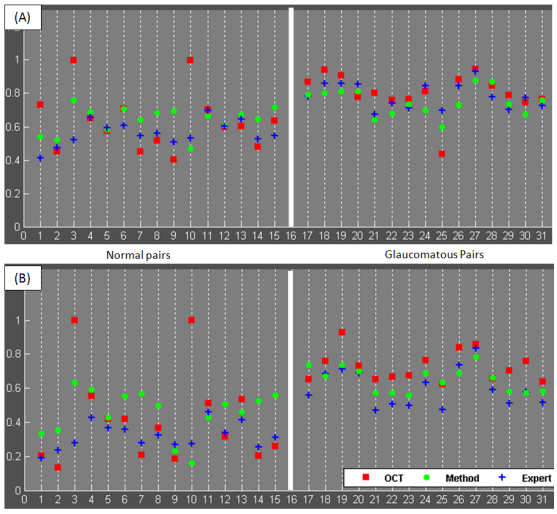

Fig. 12. A) Vertical cup-to-disk diameter ratio; B) Cup-to-disk area ratio computed on 15 normal (left set) and 15 Glaucomatous (right set) image pairs.

were computed for the quantitative analysis of the proposed method. They were compared with the same figures derived from i) the expert marking and with ii) the software provided with the OCT machine. The CDR is a widely used parameter for the glaucoma assessment and depends on the accuracy of segmentation along the vertical axis. The CAR parameter was chosen to assess the overall segmentation accuracy in all directions. Fig. 12 shows a plot of the computed values obtained for CDR (top row) and CAR (bottom row) along with the ground truth from expert marking and OCT. In both plots, the results for the 15 normal image pairs are on the left-side and the 15 glaucomatous image pairs are on the right-side. In two normal pairs, OCT software gave CDR and CAR to be 1 due to a failure in detecting cup location from the depth data.

The plots show that the proposed method gives better accuracy in CDR (rather than CAR) estimation. This can be seen by noting the distance between the green and red (OCT) data points and similarly between green and blue (expert) data points. This implies that localisation is more accurate in vertical direction compared to other directions (nasal and temporal side). The method also appears to be more accurate in CDR and CAR estimation for glaucoma cases compared to the normal cases. This is due to the fact that in glaucomatous

TABLE I
AVERAGE F-SCORE FOR PIXEL-LEVEL SEGMENTATION PERFORMANCE

	Expert-1	Expert-2	Expert-3	Expert-4
Monocular [2]	0.80	0.83	0.81	0.84
v-1	0.81	0.83	0.81	0.84
v-2	0.81	0.84	0.83	0.85

TABLE II
AVERAGE BOUNDARY DISTANCE IN THE RADIAL DIRECTION (IN PIXELS).

	Expert-1	Expert-2	Expert-3	Expert-4
Monocular [2]	27.72	24.72	24.93	23.19
v-1	27.71	24.11	26.10	23.30
v-2	26.05	23.51	22.59	21.16

cases, the cup has a stronger depth edge (due to the loss of underlying nerve fiber layer) and hence is better localised by the proposed method. In contrast, for normal cases, the cup boundary is marked by a fuzzier and weaker depth edge which is a challenge for localisation. Overall, the distance between the blue (expert) and red (OCT) data points appears to be smaller. This indicates that experts are able to estimate the cup boundary in 2D images quite well by integrating 2D visual cues such as vessel-bend and intensity changes, with the structural understanding gained from their clinical experience. However, difference in the absolute values obtained from the depth data and expert marking on the 2D projected data suggests the attainable lower limit on the error for our solution. In the next subsection, we present similar analysis on a larger dataset, Dataset-3.

3) Dataset-3: Two variants of the proposed solution were evaluated on this dataset. The cup segmentation method presented in [2] was used to derive the OD boundary while the two variants were used to find the cup boundary. Two types of quantitative analysis were carried out to assess the accuracy a) in segmenting the cup region and b) in estimating CDR and CAR parameters. The first assessment is of the cup segmentation accuracy using two measures defined in [2]: i) distance between the computed and expert marked boundaries and ii) pixel-wise segmentation accuracy using F-score.

Table I and II show the obtained F-score and boundary distance values, respectively. The results show that with respect to monocular version, there is improvement in results only for v-2 and not v-1. In contrast to v-2, v-1 considers motion cues gathered from both vessel and non-vessel regions. Therefore, for an angular sector, selection of the desirable boundary point from a larger candidate set becomes difficult. Due to pruning step of the candidate set with a vessel map in v-2, the final candidate set is sparser. The better performance of v-2 indicates the positive role of vessel detection plays and in particular of relevant vessel bends along the cup boundary. In both variants, the motion cues due to vessels present in the inner cup region was suppressed by using the pallor region masking as explained in Section III-C.

The second assessment is of the CDR and CAR parameters against expert markings. Table III and IV shows the mean μ and standard deviation σ of the error in estimating both parameters against experts. With respect to the monocular version, v-1 results in improved CDR estimation for normal

TABLE III
ESTIMATION ERROR IN CUP-TO-DISK RATIO (VERTICAL DIRECTION)

	Normal (33 sets)						Glaucoma (105 sets)						Total (138 sets)					
	Mono [2]		v-1		v-2		Mono [2]		v-1		v-2		Mono [2]		v-1		v-2	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Expert-1	0.26	0.15	0.23	0.17	0.21	0.16	0.09	0.08	0.10	0.07	0.10	0.07	0.13	0.13	0.13	0.12	0.13	0.11
Expert-2	0.18	0.11	0.14	0.11	0.13	0.11	0.10	0.08	0.11	0.08	0.11	0.08	0.12	0.09	0.12	0.09	0.12	0.09
Expert-3	0.21	0.11	0.17	0.13	0.16	0.13	0.10	0.08	0.08	0.06	0.07	0.05	0.12	0.10	0.10	0.09	0.09	0.09
Expert-4	0.21	0.13	0.18	0.14	0.17	0.13	0.09	0.08	0.09	0.06	0.09	0.06	0.12	0.11	0.11	0.09	0.10	0.09

TABLE IV
ESTIMATION ERROR IN CUP-TO-DISK AREA

	Normal (33 sets)						Glaucoma (105 sets)						Total (138 sets)					
	Mono [2]		v-1		v-2		Mono [2]		v-1		v-2		Mono [2]		v-1		v-2	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Expert-1	0.33	0.20	0.35	0.21	0.30	0.21	0.12	0.11	0.10	0.08	0.10	0.08	0.17	0.16	0.16	0.16	0.15	0.15
Expert-2	0.25	0.15	0.26	0.17	0.22	0.17	0.12	0.10	0.12	0.08	0.12	0.08	0.15	0.13	0.15	0.12	0.14	0.12
Expert-3	0.29	0.17	0.31	0.18	0.26	0.18	0.12	0.11	0.13	0.10	0.10	0.08	0.16	0.14	0.18	0.14	0.14	0.13
Expert-4	0.28	0.17	0.30	0.19	0.26	0.18	0.10	0.09	0.10	0.07	0.08	0.06	0.15	0.14	0.15	0.14	0.13	0.13

which indicates better boundary localisation accuracy in the vertical direction. Whereas for CAR, there is no gain in the performances as concluded in the first assessment. There is an improvement of 16% and 13% in both CDR and CAR estimation respectively, using v-2. This is consistent with the results of the first assessment. Overall, v-2 yields better performance improvement in normal eyes. This is due to the fact that in a normal eye, the depth discontinuity is weak at the cup boundary. This makes the selection of boundary points difficult in v-1. Constraining the candidates to be coincident with vessels aids in selecting true boundary points from a sparse set of locations and in achieving higher specificity in the detection of glaucoma.

4) Dataset-4: Based on the above performance of v-2, it was further assessed on the HRA dataset which has a fixed degree of camera motion. Sample images shown in Fig. 13 exhibit a higher contrast for OD boundary, compared to the structures present inside the disk region. The detection results, shown on magnified regions of interest in the last row, indicate successful localisation of depth edges associated with the cup boundary. Hence it can be concluded that the proposed method is also applicable to image pair acquired with fixed degree camera motion.

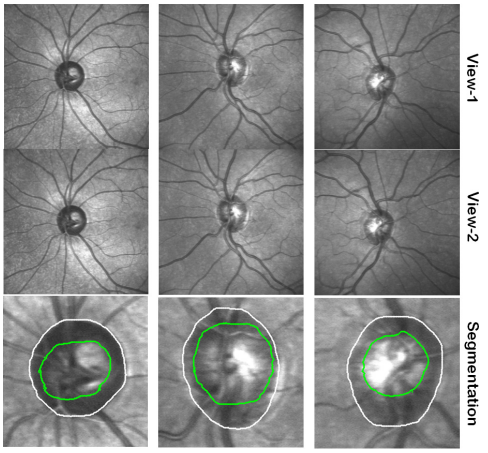

Fig. 13. Sample image pairs from HRA dataset (top 2 rows), and detected OD (white) and cup (green) boundaries (last row).

V. DISCUSSION AND CONCLUSION

In this work, a novel formulation for the problem of cup segmentation from two retinal views/images was presented. The proposed solution derives the cup boundary based on various evidences for depth discontinuities. The method is applicable to a range of image pairs acquired sequentially. Four different datasets have been used to assess the proposed method. The experimental results show the promise in the formulation of cup boundary as points of depth discontinuity along the cup boundary and as an alternative approach for cup segmentation.

The experimental results indicate that there is sufficient information encoded in the obtained evidences to infer the cup boundary. This has been shown by using multiple circular contours derived from the evidence maps. However, the final segmentation requires a selection process which integrates evidences and localises the cup boundary points. As mentioned earlier, the use of multiple views (more than 2) can provide adequately distributed, reliable information about boundary points.

In order to validate this fact, we collected three views (one additional view) for 8 glaucomatous and 7 normal eyes. The markings for the OD and cup region were obtained from a glaucoma expert. A simple extension was made in the proposed method to incorporate three views. Two sets of evidence maps were first obtained by performing pair-wise registration between reference and other two views. An example of the obtained sets is shown in Fig. 14. An additional set as a result of third view, enriches information about the cup boundary either by increasing confidence of evidences at certain locations or by providing evidences at new locations along the cup boundary. Next, average evidence maps u_x, u_y, u_o, and e_m were obtained by integrating the two sets. A procedure which is similar to the proposed one was later used to obtain cup boundary from the evidence maps. The error figure (μ/σ) in the estimation of CDR for 3 views was 0.09/0.06 and for 2 views was 0.13/0.06. In the estimation of CAR, the error figure was 0.13/0.06 for 3 views and 0.16/

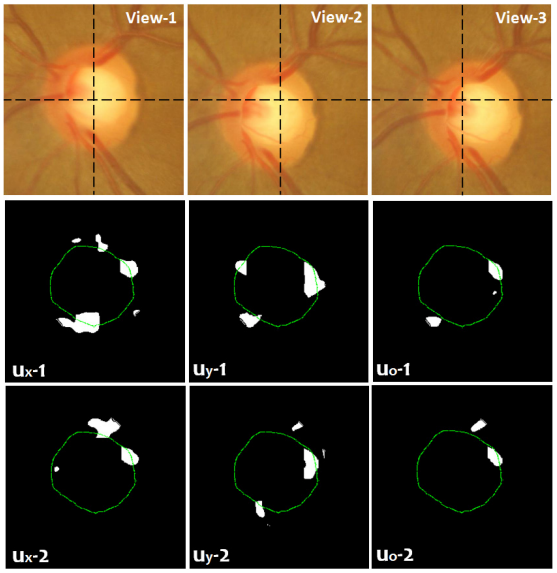

Fig. 14. Uniform and better distribution of evidences with multiple views (more than 2). First row shows three differently acquired views of the optic disk region. Second and third rows show two sets of evidence maps obtained by pairing reference view with other two views.

0.12 for 2 views. It is evident that the performance gain by the proposed solution with an additional view is significant with error reductions of 33% and 18% in the estimation of CDR and CAR, respectively. Since CDR is the parameter of choice for clinicians, significant improvement in that figure is encouraging. Further improvements can be expected with a refinement of the selection process. One option is to integrate explicitly computed vessel bend information to boost the confidence at true boundary location. The localisation of boundary points can also be enhanced by incorporating local motion boundary model [19] [28]. This model can be derived either in a supervised or unsupervised manner from the motion characteristics of the cup boundary points. The confidence assignment can be improved with an optimal assignment of weights to the individual evidence.

In summary, the proposed formulation for cup segmentation comes with following advantages: a) it relaxes a strict stereo image pair requirement and helps extend the scope of the problem, b) it shows promise in handling inconsistencies in image pairs due to varying degree and direction of camera motion; and c) it gives flexibility to integrate information from multiple views (more than two) to improve performance.

REFERENCES

[1] G. Michelson, S. Wrntges, J. Hornegger, and B. Lausen, "The papilla as screening parameter for early diagnosis of glaucoma," *Deutsches Aerzteblatt international*, vol. 105, pp. 34–35, 2008.

[2] G. D. Joshi, J. Sivaswamy, and S. R. Krishnadas, "Optic disk and cup segmentation from monocular colour retinal images for glaucoma assessment," *IEEE Transactions on Medical Imaging*, vol. 30, no. 6, 2011.

[3] J. Xu, O. Chutatape, E. Sung, C. Zheng, and P. Chew, "Optic disk feature extraction via modified deformable model technique for glaucoma analysis," *Pattern Recognition*, vol. 40, no. 7, pp. 2063–2076, 2007.

[4] D. Wong, J. Liu, J. H. Lim, H. Li, X. Jia, F. Yin, and T. Wong, "Automated detection of kinks from blood vessels for optic cup segmentation in retinal images," *Proc. SPIE, Medical Imaging*, p. 72601J, 2009.

[5] G. D. Joshi, J. Sivaswamy, K. Karan, P. R., and R. Krishnadas, "Vessel bend-based cup segmentation in retinal images," *Proc. Int. Conf. Pattern Recognition (ICPR)*, pp. 2536–2539, 2010.

[6] G. D. Joshi, J. Sivaswamy, K. Karan, and R. Krishnadas, "Optic disk and cup boundary detection using regional information," *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 948–951, 2010.

[7] J. Liu, D. Wong, J. Lim, H. Li, N. Tan, and T. Wong, "Argali- an automatic cup-to-disc ratio measurement system for glaucoma detection and analysis framework," *Proc. SPIE, Medical Imaging*, pp. 72 603K–8, 2009.

[8] M. D. Abramoff, W. L. M. Alward, E. C. Greenlee, L. Shuba, C. Y. Kim, J. H. Fingert, and Y. H. Kwon, "Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features," *Investigative Ophthalmology and Visual Science*, vol. 48, pp. 1665–1673, 2007.

[9] M. B. Merickel, X. Wu, M. Sonka, and M. Abramoff, "Optimal segmentation of the optic nerve head from stereo retinal images," *Proc. SPIE: Medical Imaging*, vol. 6143, pp. 1031–1038, 2006.

[10] E. Corona, S. Mitra, M. Wilson, T. Krile, and Y. H. K. P. Soliz, "Digital stereo image analyzer for generating automated 3-d measures of optic disc deformation in glaucoma." *IEEE Trans Med Imaging.*, vol. 21, no. 10, pp. 1244–1253, 2002.

[11] A. Guesalag, P. Irarrzabal, M. Guarini, and R. lvarez, "Measurement of the glaucomatous cup using sequentially acquired stereoscopic images," *Measurement*, vol. 34, no. 3, pp. 207–213, 2003.

[12] C. Muramatsu, T. Nakagawa, A. Sawada, Y. Hatanaka, T. Hara, T. Yamamoto, and H. Fujita, "Determination of cup and disc ratio of optical nerve head for diagnosis of glaucoma on stereo retinal fundus image pairs," *Proc. SPIE. Medical Imaging,*, pp. 72 603L–2, 2009.

[13] J. A. Lusk, "Modeling 3-d reconstruction by image rectification of stereo images acquired by cameras of unknown and varying parameters," *MS Thesis: Texas Tech University*, 2007.

[14] J. Xu and O. Chutatape, "Auto-adjusted 3-d optic disk viewing from low-resolution stereo fundus image," *Comput Biol Med.*, vol. 36, no. 9, pp. 921–940, 2006.

[15] S. K. Kim, J. M. Seo, H. J. Kong, H. Chung, K. H. Park, J. M. Hwang, K. S. Park, and H. C. Kim, "3d reconstruction from stereo disc photograph based on block matching method with variable search range," *Proc. IFMBE*, vol. 11, ISSN: 1727-1983, no. 1, 2005.

[16] R. S. Feris, "Detection and modeling of depth discontinuities with lighting and viewpoint variation," *Ph.D Thesis: University of California, Santa Barbara*, Sep-2006.

[17] S. Birchfield, "Depth and motion discontinuities," *PhD thesis, Stanford University*, 1999.

[18] D. Lee, "Coping with discontinuities in computer vision: Their detection, classification, and measurement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 4, pp. 321–344, 1999.

[19] M. J. Black and P. Anandan, "Constraints for the early detection of discontinuity from motion," *Proc. AAAI*, vol. 2, pp. 1060–1066, 1990.

[20] A. Spoerri and S. Ullman, "The early detection of motion boundaries," *Proc. ICCV*, pp. 209–218, 1987.

[21] K. M. Mutch and W. B. Thompson, "Analysis of accretion and deletion at boundaries in dynamic scenes," *IEEE Trans. Pattern Analysis and Machine Learning*, vol. 7, no. 2, p. 133138, 1985.

[22] J. Thirion, "Image matching as a diffusion process: an analogy with maxwells demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, 1998.

[23] D.-J. Kroon and C. H. Slump, "Mri modalitiy transformation in demon registration," *Proc. IEEE Int. Symp. Biomedical Imaging*, pp. 963–966, 2009.

[24] H. Wang, L. Dong, J. ODaniel, R. Mohan, K. K. A. A. S. Garden, D. A. Kuban, J. Y. Bonnen, M. Chang, and R. Cheung, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," *Physics in Medicine and Biology*, vol. 50(12), pp. 2887–2905, 2005.

[25] T. Vercauterena, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2008.

[26] S. Garg, J. Sivaswamy, and S. Chandra, "Unsupervised curvature-based retinal vessel segmentation," *Proc. ISBI*, pp. 344–347, 2007.

[27] L. Giancardo, F. Meriaudeau, T. P. Karnowski, K. W. Tobin, E. Grisan, P. Favaro, A. Ruggeri, and E. Chaum, "Textureless macula swelling detection with multiple retinal fundus images," *IEEE Trans Biomed Eng.*, vol. 58, no. 3, pp. 795–799, 2011.

[28] H. Liu, T. hong Hong, M. Herman, and R. Chellappa, "Motion-model-based boundary extraction," *University of Maryland*, 1995.