Enhanced video indirect ophthalmoscopy (VIO) via robust mosaicing

Rolando Estrada,[1,*] Carlo Tomasi,[1] Michelle T. Cabrera,[2] David K. Wallace,[2] Sharon F. Freedman,[2] and Sina Farsiu[2,3]

[1]*Dept. of Computer Science, Duke University, Durham, NC 27708*
[2]*Dept. of Ophthalmology, Duke University, Durham, NC 27708*
[3]*Dept. of Biomedical Engineering, Duke University, Durham, NC 27708*

[*]*restrada@cs.duke.edu*

Abstract: Indirect ophthalmoscopy (IO) is the standard of care for evaluation of the neonatal retina. When recorded on video from a head-mounted camera, IO images have low quality and narrow Field of View (FOV). We present an image fusion methodology for converting a video IO recording into a single, high quality, wide-FOV mosaic that seamlessly blends the best frames in the video. To this end, we have developed fast and robust algorithms for automatic evaluation of video quality, artifact detection and removal, vessel mapping, registration, and multi-frame image fusion. Our experiments show the effectiveness of the proposed methods.

OCIS codes: (100.0100) Image processing; (100.2960) Image analysis; (170.4470) Ophthalmology.

References and links

1. B. Zitova and J. Flusser, "Image registration methods: a survey," Image Vision Comput. **21**, 977–1000 (2003).
2. M. D. Robinson, C. A. Toth, J. Y. Lo, and S. Farsiu, "Efficient Fourier-wavelet super-resolution," IEEE Trans. Image Process. **19**, 2669–2681 (2010).
3. W. Tasman, A. Patz, J. McNamara, R. Kaiser, M. Trese, and B. Smith, "Retinopathy of prematurity: the life of a lifetime disease," Am. J. Ophthalmol. **141**, 167–174 (2006).
4. D. K. Wallace, G. E. Quinn, S. F. Freedman, and M. F. Chiang, "Agreement among pediatric ophthalmologists in diagnosing plus and pre-plus disease in retinopathy of prematurity," J. Am. Assoc. Pediatric Opthalmol. Strabismus **12**, 352–356 (2008).
5. R. Gelman, M. Martinez-Perez, D. Vanderveen, A. Moskowitz, and A. Fulton, "Diagnosis of plus disease in retinopathy of prematurity using Retinal Image multiScale Analysis," Invest. Ophthalmol. & Visual Sci. **46**, 4734–4738 (2005).
6. D. Wallace, Z. Zhao, and S. Freedman, "A pilot study using" ROPtool" to quantify plus disease in retinopathy of prematurity." J. Am. Assoc. Pediatric Opthalmol. Strabismus **11**, 381–387 (2007).
7. S. Ahmad, D. Wallace, S. Freedman, and Z. Zhao, "Computer-assisted assessment of plus disease in retinopathy of prematurity using video indirect ophthalmoscopy images," Retina **28**, 1458–1462 (2008).
8. F. Zana and J. Klein, "A multimodal registration algorithm of eye fundus images using vessels detection and hough transform," IEEE Trans. Med. Imaging **18**, 419–428 (1999).
9. C. Stewart, C. Tsai, and B. Roysam, "The dual-bootstrap iterative closest point algorithm with application to retinal image registration," IEEE Trans. Med. Imaging **22**, 1379–1394 (2003).
10. R. González and R. Woods, *Digital Image Processing*, 3rd ed.(Prentice Hall, 2008).
11. H. Cheng, X. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," Pattern Recognition **34**, 2259–2281 (2001).
12. J. Kautsky, J. Flusser, and B. Zitová, "A new wavelet-based measure of image focus," Pattern Recognition Lett. **23**, 1785–1794 (2002).

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

13. M. Subbarao and J. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," IEEE Trans. Pattern Anal. Machine Intell. **20**, 864–870 (1998).
14. S. Fox, R. Silver, E. Kornegay, and M. Dagenais, "Focus and edge detection algorithms and their relevance to the development of an optical overlay calibration standard," Proc. SPIE **3677**, 95-106 (1999).
15. Y. Zhang, Y. Zhang, and C. Wen, "A new focus measure method using moments," Image Vision Comput. **18**, 959–965 (2000).
16. D. Tsai and C. Chou, "A fast focus measure for video display inspection," Machine Vision Appl. **14**, 192–196 (2003).
17. C. Sun, "De-interlacing of video images using a shortest path technique," IEEE Trans. Consumer Electron. **47**, 225–230 (2001).
18. S. Lin, Y. Chang, and L. Chen, "Motion adaptive interpolation with horizontal motion detection for deinterlacing," IEEE Trans. Consumer Electron. **49**, 1256–1265 (2003).
19. D. Tschumperlé and B. Besserer, "High quality deinterlacing using inpainting and shutter-model directed temporal interpolation," Comput. Vision Graphics **32**, 301–307 (2006).
20. D. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of International Conference on Computer Vision* (IEEE, 1999), pp. 1150–1157.
21. P. Burt and E. Adelson, "A multiresolution spline with application to image mosaics," ACM Trans. Graphics **2**, 217–236 (1983).
22. E. Peli, "Contrast in complex images," J. Opt. Soc. Am. A **7**, 2032–2040 (1990).
23. J. Staal, M. Abràmoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," IEEE Trans. Med. Imaging **23**, 501–509 (2004).
24. J. Soares, J. Leandro, R. Cesar Jr, H. Jelinek, and M. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," IEEE Trans. Med. Imaging **25**, 1214–1222 (2006).
25. J. Starck, F. Murtagh, E. Candes, and D. Donoho, "Gray and color image contrast enhancement by the curvelet transform," IEEE Trans. Image Processing **12**, 706–717 (2003).
26. T. Acharya and A. Ray, *Image Processing: Principles and Applications* (Wiley-Interscience, 2005).
27. J. Movellan, "Tutorial on Gabor filters," Open Source Document (2002), http://mplab.ucsd.edu/wordpress/tutorials/gabor.pdf.
28. S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," IEEE Trans. Med. Imaging **8**, 263–269 (1989).
29. Q. Li, J. You, L. Zhang, and P. Bhattacharya, "Automated retinal vessel segmentation using Gabor filters and scale multiplication," in *Proceedings International Conference on Image Processing, Computer Vision, and Pattern Recognition* (WORLDCOMP, 2006), pp. 3521–3527.
30. C. Wilson, K. Cocker, M. Moseley, C. Paterson, S. Clay, W. Schulenburg, M. Mills, A. Ells, K. Parker, G. Quinn, A. Fielder, and J. Ng, "Computerized analysis of retinal vessel width and tortuosity in premature infants," Invest. Ophthalmol. & Visual Sci. **49**, 3577–3585 (2008).
31. C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," ACM Comput. Surv. **36**, 81–121 (2004).
32. R. Szeliski, "Image alignment and stitching: A tutorial," Found. Trends Comput. Graphics Vision **2**, 1–104 (2006).
33. A. Zomet, A. Levin, S. Peleg, and Y. Weiss, "Seamless image stitching by minimizing false edges," IEEE Trans. Image Process. **15**, 969–977 (2006).
34. L. Brown, "A survey of image registration techniques," ACM Comput. Surv. **24**, 325–376 (1992).
35. J. Maintz and M. Viergever, "A survey of medical image registration," Med. Image Anal. **2**, 1–36 (1998).
36. M. D. Robinson, S. Farsiu, and P. Milanfar, "Optimal registation of aliased images using variable projection with applications to super-resolution," Comput. J. **52**, 31–42 (2009).
37. K. Simonson, S. Drescher, and F. Tanner, "A statistics-based approach to binary image registration with uncertainty analysis," IEEE Trans. Pattern Anal. Machine Intell. **29**, 112–125 (2007).
38. P. Besl and N. McKay, "A method for registration of 3-d shapes," IEEE Trans. Pattern Anal. Machine Intell. **14**, 239–256 (1992).
39. S. Weik, "Registration of 3-d partial surface models using luminance and depth information," in *Proceedings of the International Conference on Recent Advances in 3-D Digital Imaging and Modeling* (IEEE Computer Society, 1997), pp. 93–100.
40. A. Wong, W. Bishop, and J. Orchard, "Efficient multi-modal least-squares alignment of medical images using quasi-orientation maps," in *Proceedings of International Conference on Image Processing, Computer Vision, and Pattern Recognition* (WORLDCOMP, 2006), pp. 74–80.
41. A. Fitch, A. Kadyrov, W. Christmas, and J. Kittler, "Fast robust correlation," IEEE Trans. Image Process. **14**, 1063–1073 (2005).
42. B. Srinivasa and B. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," IEEE Trans. Image Process. **5**, 1266–1271 (1996).
43. G. Wolberg and S. Zokai, "Robust image registration using log-polar transform," in *Proceedings of IEEE International Conference on Image Processing* (IEEE, 2000), pp. 493–496.

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

44. J. Shi and C. Tomasi, "Good features to track," in *Proceedings of IEEE Computer Vision and Pattern Recognition* (IEEE, 1994), pp. 593–593.
45. M. Daszykowski, K. Kaczmarek, Y. Vander Heyden, and B. Walczak, "Robust statistics in data analysis–a review:: basic concepts," Chemometrics Intell. Lab. Syst. **85**, 203–219 (2007).
46. S. Farsiu, M. Elad, and P. Milanfar, "Constrained, globally optimal, multi-frame motion estimation," in *Proceedings of IEEE Workshop on Statistical Signal Processing* (IEEE, 2005), pp. 1396–1401.
47. R. Fletcher, "Conjugate gradient methods for indefinite systems," Numerical Anal. **506**, 73–89 (1976).
48. H. Van der Vorst, "Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems," SIAM J. Sci. Statist. Comput. **13**, 631–644 (1992).
49. D. Zwillinger, *CRC Standard Mathematical Tables and Formulae* (Chapman & Hall/CRC, 2003).
50. J. Phillips, "Survey of absolute orientation techniques," http://www.cs.duke.edu/jeffp/triseminar/absolute-orientation.pdf.
51. B. Horn, "Closed-form solution of absolute orientation using unit quaternions," J. Opt. Soc. Am. A **4**, 629–642 (1987).
52. Y. K. Tao, S. Farsiu, and J. A. Izatt, "Interlaced spectrally encoded confocal scanning laser ophthalmoscopy and spectral domain optical coherence tomography," Biomed. Opt. Express **1**, 431–440 (2010).

1. Introduction

Hardware improvements yield quickly diminishing returns when gathering useful information from medical images, because the optical components necessary to capture very *high-quality* scans become prohibitively expensive for many practical applications. The image processing community has developed several multi-frame image fusion algorithms [1, 2] that generate high-quality imagery from lower-quality imaging detectors. In this paper, we develop fast and robust multi-frame image fusion algorithms to produce wide Field of View (FOV) and artifact-free images from a large collection of narrow FOV images of varying quality.

While the proposed algorithms are general and can be adapted to a variety of image enhancement and analysis applications, this paper targets a very challenging medical imaging scenario. We address the problem of generating high-quality images from non-sedated premature infants' eyes captured during routine clinical evaluations of the severity of Retinopathy of Prematurity (ROP). ROP is disorder of the retinal blood vessels which is a major cause of vision loss in premature neonates, in spite of being preventable with timely treatment [3]. Important features of the disease include increased diameter (dilation) as well as increased tortuosity (wiggliness) of the retinal blood vessels in the portion of the retina centered on the optic nerve (the posterior pole). Studies have shown that when the blood vessels in the posterior pole show increased dilation and tortuosity (called pre-plus in intermediate, and plus in severe circumstances), this correlates well with the severity of the ROP [4].

Thus, an important prognostic sign of severe ROP is the presence of plus disease, consisting of dilation and tortuosity of retinal vessels. Plus disease is the primary factor in determining whether an infant with ROP requires laser treatment. Unfortunately, human assessment of plus disease is subjective and error-prone. A previous study showed that ophthalmologists disagree on the presence or absence of plus disease in 40% of retinal images [4].

Semi-automated image analysis tools such as ROPtool [4] and RISA [5] show similar or even superior sensitivity and specificity compared to individual pediatric ophthalmologists when high-quality retinal photographs are obtained with the RetCam imaging system (Clarity Medical Systems, Inc., Pleasanton, CA). The full details of the procedure for using ROPTool have been previously published [6]. In summary, ROPTool displays the image to be analyzed and the operator identifies the key anatomical parts of the retina, such as the optic nerve and the vessels in each quadrant, by clicking on the image. However, RetCam is expensive and inconvenient for imaging pediatric patients, and is not commonly used during routine examinations. Instead, examination with the Indirect Ophthalmoscope (IO) is the standard of care for ROP evaluation of neonate eyes. A Video Indirect Ophthalmoscope (VIO) is a relatively inexpensive imaging system (about 6 times cheaper than RetCam) and much more convenient for capturing retinal

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

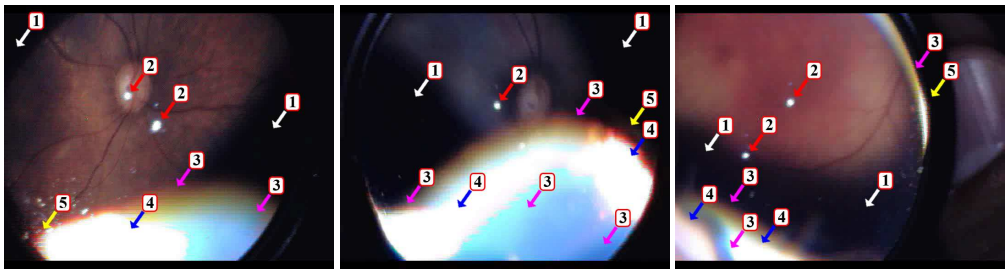

Fig. 1. VIO frame artifacts: Three sample VIO frames, from three different videos, display a number of common artifacts. Each arrow's number and color indicate the type of artifact: **(1)** (white) black regions; **(2)** (red) white spots; **(3)** (magenta) artificial colors; **(4)** (blue) saturation near the lens' rim; **(5)** (yellow) interlacing artifacts. All but the interlacing artifacts are produced by the optics of the hand-held condensing lens.

images during IO examinations. In VIO, the physician wears a head-mounted video camera during routine IO evaluations. Unfortunately, many individual VIO frames have poor quality, and a previous study reports that only 24% of these videos can be utilized for ROP evaluation with ROPtool [7].

Several types of artifacts make the raw recorded VIO images difficult to analyze automatically, including interlacing artifacts, brightness saturation, white or black spots, and distorted colors. Frames often contain non-retinal objects, as shown in Fig. 4 (a), and have a narrow FOV. Some of these problems are highlighted in Fig. 1. Furthermore, a raw VIO video indiscriminately records every part of an IO examination, and therefore contains numerous spurious and low quality frames that need to be removed prior to any form of automated analysis.

In this paper, we develop a framework for obtaining the relevant retinal data from a VIO video in the form of a single, high quality image, suitable for analysis manually or with semi-automated tools such as ROPTool. While multi-fundus image registration methods exist, such as [8, 9], the VIO frames' low quality and large number of artifacts and spurious objects adversely affects the performance of these methods. Our proposed video processing pipeline (Fig. 2) involves novel algorithms for: (1) rapid detection of the most relevant and highest-quality VIO images, (2) detection and removal of VIO imaging artifacts, (3) extraction and enhancement of retinal vessels, (4) registration of images with large non-translational displacement under possibly varying illumination, and (5) seamless image fusion. We validated the diagnostic usability of our technique for semi-automated analysis of plus disease by testing how well the semi-automated diagnosis obtained using our mosaics matched an expert physician's diagnosis.

The rest of this paper is organized as follows: We select frames in Section 2. We then enhance the image by removing artifacts in Section 3, and map vessels in Section 4. In Section 5, we fuse the enhanced images through frame registration, color mapping, and pixel selection. We present our experimental results in Section 6 and discuss future research and clinical applications in Section 7.

2. Frame selection

Only a small fraction of the frames in raw VIO data is suitable for analysis: patient preparation and switching from eye to eye result in numerous frames that do not feature the retina. In addition, as Fig. 1 illustrates, a significant portion of retinal frames are poorly focused and are marred by artifacts that arise from the condensing lens, sensor noise, and video compression. Manually searching for the best frames in a VIO sequence is impractical. In our approach, we find retinal frames by the percentage of pixels with retina-like colors, and we estimate the degree of image focus by the ratio of the signal contents at intermediate and high bandpass frequencies. We explain these two criteria in the following subsections.

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

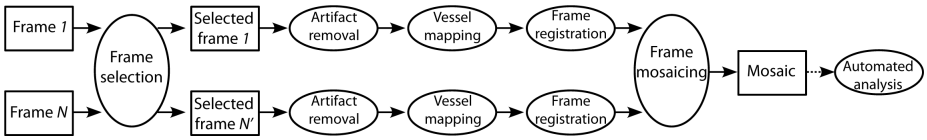

Fig. 2. Retinal mosaicing pipeline: Our proposed pipeline generates a single, high quality mosaic from a raw VIO recording. We select frames based on a hue-saturation-value (HSV) quality score, a spatial frequency measure, and a combination of the two measures. We then remove artifacts present in the selected frames and compute a high contrast vessel map from each of them. We register the selected frames based on these maps and fuse the registered frames into a single mosaic, suitable for human and semi-automated analysis.

In our framework, we view each frame as an $X \times Y \times 3$ matrix I. The indexed set of N frames is denoted as $\mathbf{I} = I_1, I_2, ..., I_N$. A pixel position in I is given by a two-dimensional vector of integers, $\mathbf{p} = [x, y]^{\mathrm{T}}$. The set of valid pixel positions, such that $x \in [1, X]$ and $y \in [1, Y]$, is denoted by P. The value at each pixel position for a given frame is given by a 3-dimensional vector $\mathbf{v}$:

$$I(\mathbf{p}) = \mathbf{v}. \tag{1}$$

Unless otherwise specified, $\mathbf{v}$ represents an RGB value with entries normalized between 0 and 1.

2.1. HSV classification

We classify pixels in the Hue-Saturation-Value (HSV) color space [10] to assign a retinal or non-retinal label to every pixel. Classification in HSV space is more robust to highlights, shadows, and texture variations than in other color spaces [11]. We use color for pixel classification because we empirically determined that retinal pixels for a given patient exhibit a narrow and consistent color distribution. Furthermore, the color distribution of any frame can be determined reliably and efficiently.

In HSV classification, we specify a closed decision boundary S in HSV space. Vectors within S are classified as retinal pixels, while those outside S are labeled non-retinal. We specify S as the Cartesian product of three intervals, S_H, S_S, and S_V. We eschew more complex boundaries for computational efficiency. We construct the set P_R of retinal pixels as those whose HSV values lie within S. A frame's HSV score is given by the proportion of retinal pixels in the image:

$$h(I) = \frac{|P_R|}{|P|}. \tag{2}$$

where the bars denote set cardinality. Fig. 3 shows a sample color distribution, and Fig. 4 (a) displays a series of frames from a single video ranked by their HSV score, from highest to lowest. For further processing, we retain only a fixed number of top-score frames.

2.2. Spatial frequency estimation

Only a fraction of VIO frames are properly focused. Camera and patient motion lead to blurry frames and frames with interlacing artifacts, such as those in Fig. 1, both of which need to be discarded prior to image fusion. HSV classification accurately identifies frames with high retinal content, but does not account for these types of image degradation. We further refine the selected frame sequence I' by determining a spatial frequency score for each frame.

A number of approaches have tackled image focus [12, 13, 14, 15, 16]. Most focusing methods measure the quantity of high frequencies in an image using image gradients [13] or high-pass filtering [12].

For our application, however, the presence of high frequency interlacing artifacts [17, 18] precludes the direct use of these techniques. When the scene motion is fast relative to the frame

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

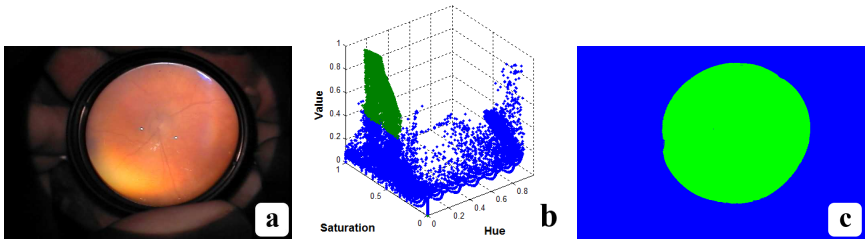

Fig. 3. Frame HSV color distribution: (a) A sample VIO frame. (b) The scatter plot of the HSV color values of the sample frame. Retinal pixels (green), exhibit a narrow color distribution in HSV space relative to the rest of the image (blue). While the retinal pixels constitute 30% of the image, they are more tightly clustered than the non-retinal pixels. (c) Color-coded frame: retinal pixels are shown in green and non-retinal pixels in blue.

capture rate in interlaced video, a combing effect arises [19]. This interlacing artifact occurs across adjacent rows of a single frame, so it has very high frequency.

To account for the simultaneous presence of interlacing and blurring, we observe that relevant anatomical information in the frames is present at intermediate spatial frequencies. Very high frequencies correspond to interlacing artifacts, while blurry images only contain low frequency values. We employ differences of Gaussians [20] in the frequency domain to obtain two band-pass filters at high and intermediate frequencies. We then estimate a ratio of the norms of the intermediate and high band-pass frequencies. A high ratio reveals sharp frames with little interlacing. The use of a ratio between higher and lower pass bands was shown by Kautsky et al. [12] to be both robust and monotonic with respect to the degree of defocus. Unlike their original ratio, however, we account for high frequency interlacing by using a band-pass, rather than a low-pass image.

To obtain the band-pass filters, we first construct two low-pass Gaussian filters $\mathscr{N}_{\sigma_m}$ and $\mathscr{N}_{\sigma_l}$:

$$\mathscr{N}_{\sigma_m} = \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-[(u-\mu_u)^2/2\sigma_m^2 + (v-\mu_v)^2/\sigma_m^2]}, \quad \mathscr{N}_{\sigma_l} = \frac{1}{\sqrt{2\pi\sigma_l^2}} e^{-[(u-\mu_u)^2/2\sigma_l^2 + (v-\mu_v)^2/2\sigma_l^2]}. \quad (3)$$

The parameter σ controls the amount of smoothing, and we set $\sigma_m = 2\sigma_l$. Note that $\mathscr{N}_{\sigma_m}$ is anisotropic, while $\mathscr{N}_{\sigma_l}$ is isotropic. This is to account for the fact that interlacing only occurs between image rows, not columns. We then compute the Fast Fourier Transform (FFT) of each frame, $\mathbf{F}_h = \mathscr{F}\{I\}$, and window it with $\mathscr{N}_{\sigma_m}$:

$$\mathbf{F}_m(u,v) = \mathbf{F}_h(u,v)\mathscr{N}_{\sigma_m}(u,v). \quad (4)$$

$\mathscr{N}_{\sigma_m}(u,v)$ attenuates the very high frequencies of $\mathbf{F}_m$ and more so in the y direction. We then window $\mathbf{F}_m$ with the second Gaussian filter $\mathscr{N}_{\sigma_l}$:

$$\mathbf{F}_l(u,v) = \mathbf{F}_m(u,v)\mathscr{N}_{\sigma_l}(u,v). \quad (5)$$

Our spatial frequency measure is given by the ratio of the differences of the two operations:

$$b(I) = \frac{\|\mathbf{F}_m - \mathbf{F}_l\|_1}{\|\mathbf{F}_h - \mathbf{F}_m\|_1}. \quad (6)$$

The above measure is a ratio of intermediate to high band-pass frequency norms, and accounts for both interlacing artifacts and blurring: $b(I)$ will only have a high value when the frame contains significant intermediate frequencies (vessels) and few very high frequencies (interlacing artifacts). As Fig. 4 (b) shows, our spatial frequency measure gives the highest score to frames with high vessel content and minimal interlacing distortion. The successive smoothing scheme described above is similar to a Laplacian pyramid [21], but we apply the Gaussian filters in the frequency domain with no downsampling.

The cascade of selection steps based on color and spatial frequency criteria still leaves a large number of frames for mosaicing. We reduce this number further by retaining only those frames

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

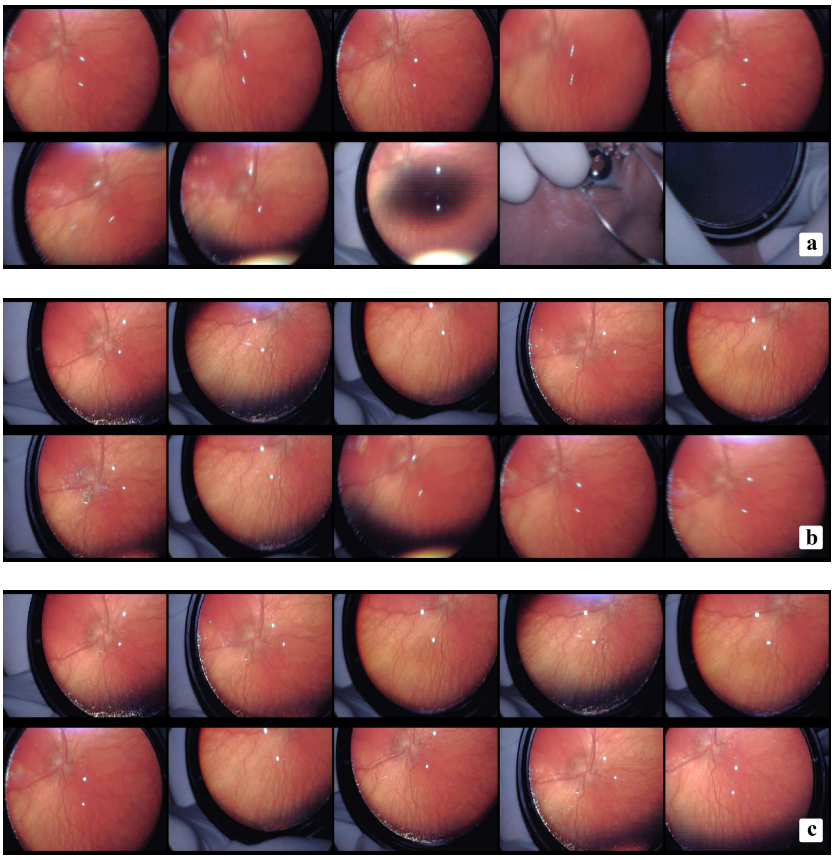

Fig. 4. Frames classified by quality scores: Frames from a 2500-frame video ranked based on the (a) HSV and (b) spatial frequency scores. Scores decrease from left to right and top to bottom. The frames in (c) are ranked by the convex combination (7) of the two scores, with $\upsilon = 0.3$. The convex combination score balances crispness with coverage. Each collage shows frames with ranks 1 through 10. Images are best viewed on screen.

with a high value for the convex combination of the normalized HSV and spatial frequency scores for each frame:

$$q(I) = \upsilon h(I) + (1-\upsilon)b(I), \quad \forall I \in \mathbf{I}'. \tag{7}$$

This combined score balances retinal coverage with vessel crispness, as Fig. 4 (c) illustrates.

3. Artifact removal

Lens and compression artifacts are present in most, if not all, raw VIO frames, as noted in Section 1 and as Fig. 1 illustrates. Furthermore, when frames are fused, each frame's artifacts accumulate in the resulting mosaic. Figure 5 shows how the existence of even a few white spots per frame can overwhelm a naive fusion of several frames.

We address these problems by directly removing artifacts and non-retinal regions from the source frames prior to further processing. To this end, we use directional local contrast filtering to remove high-saliency lens artifacts. We then fully mask non-retinal regions using HSV color classification.

Lens artifacts saturate the video's luminance, resulting in regions of high contrast with respect to the local background. We make use of this visual saliency to detect and remove these artifacts. We model saliency based on Weber's measure of contrast [22]:

$$c_o = \frac{m_o - m_b}{m_b}, \tag{8}$$

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

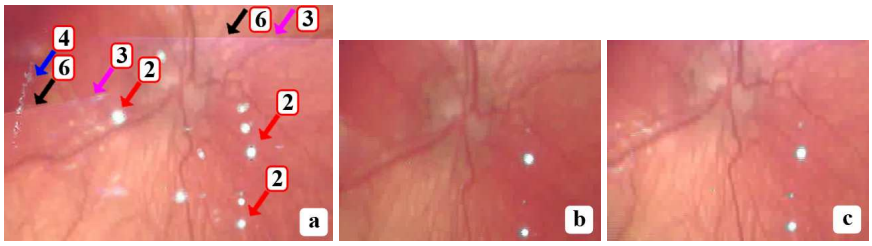

Fig. 5. Accumulating artifacts can overwhelm naive frame fusion: (a) A close-up of the naive fusion of five frames with no artifact removal. White spots, speckles, distorted colors and other artifacts from various frames accumulate in the mosaic. Following the labeling scheme of Fig. 1, the colored arrows indicate the different types of artifacts: **(2)** (red) white spots; **(3)** (magenta) distorted colors; **(4)** (blue) saturation caused by the lens' rim; **(6)** (black) spurious inter-frame borders. In (b) and (c), two of the originating frames are shown, which exhibit fewer artifacts in the same region.

where m_o is the median intensity of the object and m_b is the median intensity of the background. We use the median value, as opposed to the mean, due to its greater robustness to outliers. The Weber contrast measure is defined for grayscale images. We apply it to RGB images by considering only the green channel, as is standard practice for retinal images [23, 24] because of its stronger vessel contrast. We define a local Weber contrast measure for each individual pixel $\mathbf{p}$ by determining m_b from a small, rectangular neighborhood around $\mathbf{p}$:

$$c_{\mathbf{p}} = \frac{\mathbf{v}^g - m_b^g(\mathbf{p})}{m_b^g(\mathbf{p})}, \tag{9}$$

where the g superscript indicates the green channel. The exact value of $m_b^g(\mathbf{p})$ depends on the size of the neighborhood window. In our experiments, results were robust to variations in window size, as long as the window is larger than the targeted artifacts.

The sign of $c_{\mathbf{p}}$ is different for bright ($c_{\mathbf{p}} > 0$) and dark ($c_{\mathbf{p}} < 0$) contrast. Regions of dark contrast include the vessels. Therefore, we do not modify these pixels. The only anatomical region in VIO frames with bright contrast is the optic nerve head, which is not diagnostically relevant for ROP. We therefore identify pixels with bright contrast that exceed a threshold value t_c and replace them with the corresponding median value:

$$c_{\mathbf{p}} > t_c \Rightarrow I(\mathbf{p}) \leftarrow m_b(\mathbf{p}) \, . \tag{10}$$

Figure 6 illustrates the effects of these operations. The threshold value t_c was determined empirically in our experiments. In multi-frame fusion, bright contrast pixels can also be replaced with the corresponding pixel values from overlapping frames instead of the local median value.

3.1. Distorted color adjustment

The condensing lens of the ophthalmoscope often produces saturation near its rim, clearly visible in all three samples in Fig. 1. Pixels around the edges of these regions (see magenta arrows in Fig. 1) have their hue values distorted, particularly towards magenta, because of the nearby saturation. Directly incorporating these regions into the mosaic creates novel artifacts, as Fig. 5 shows.

We minimize the hue distortion around saturated regions by replacing the hue of these pixels with the nearest hue value that lies inside the color classification surface S, effectively enlarging the frame's retinal area. We identify the affected pixels by expanding the hue interval into $S'_H = [S_{H_{min}} - t_p, S_{H_{max}} + t_p]$, where $2t_p$ is the extent of expansion. With S'_H, we construct a new closed boundary S' in HSV space. We then determine the set of distorted color pixels by taking the difference of the sets:

$$S^d = S' \setminus S. \tag{11}$$

We adjust the pixels that belong to S^d by shifting their hue values so that they lie within S:

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

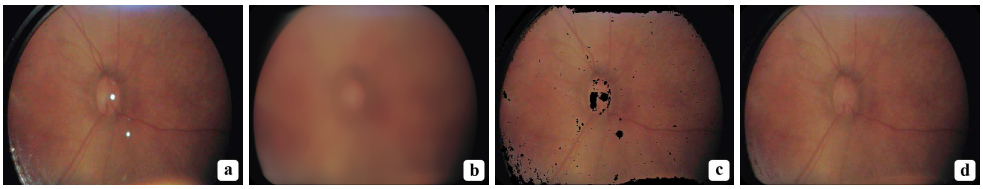

Fig. 6. The steps of directional local contrast filtering: (a) The original frame. (b) The local median for a 50×50 pixel filtering window $V_{\mathbf{p}}$. (c) Pixels that far exceed the local median brightness are marked as invalid (black in the image). (d) Invalid pixels are replaced with the local median values. This removes white spots and speckles.

$$\mathbf{v}_{\mathrm{HSV}} \in S^d \Rightarrow v^h_{\mathrm{HSV}} \leftarrow v^h_{\mathrm{HSV}} \pm t_p. \tag{12}$$

Figure 7 illustrates this adjustment. After artifact removal, we mask any pixels which fall outside the HSV boundary S to remove the remaining spurious objects such as the physician's hands, the condensing lens' disk, and other surrounding objects, as shown in Fig. 8.

4. Vessel mapping

In VIO frames, the vessels are often too faint to be detected reliably. In this section, we propose a two-step algorithm to maximize the contrast between vascular and non-vascular pixels. General contrast enhancement methods [25, 26] are insufficient to address this issue because we wish to selectively enhance only the vessels in the image and not other parts of the retina. We enhance the detectability of retinal vessels through a multi-scale approach using Laplacian-of-Gaussian (LoG) filters and Gabor wavelets [27]. Filtering kernels are widely used to enhance retinal vessels [28, 24, 29, 30, 31] due to the vessels' marked elongation and their narrow intensity distribution relative to the surrounding tissue.

4.1. LoG filter bank

The LoG filter convolves the image with a LoG operator, which is the result of convolving a low-pass Gaussian filter with a contrast-sensitive Laplace operator:

$$L_\sigma(\mathbf{p};\sigma) = (\nabla^2 \mathcal{N}(\mathbf{p};\sigma)) * I_g(\mathbf{p}), \tag{13}$$

where I_g is a grayscale version of I, "$*$" is the convolution operator, and $\mathcal{N}(\mathbf{p};\sigma)$ is an isotropic, zero-mean Gaussian kernel with variance σ^2. To enhance vessels at different scales, we convolve the original image with filters that vary in the value of their scale parameter σ and retain the maximum response at every pixel:

$$L(\mathbf{p}) = \max_\sigma L_\sigma(\mathbf{p};\sigma). \tag{14}$$

4.2. Gabor wavelet bank

To enhance vessel connectivity, we make use of Gabor wavelets, defined by multiplying a complex sinusoid by an Gaussian kernel [27]:

$$B(\mathbf{p};\lambda,\Sigma,\theta) = s(\mathbf{p};\lambda)\mathcal{N}'(\mathbf{p};\Sigma,\theta), \tag{15}$$

where $s(\cdot)$ is the sinusoidal component and $\mathcal{N}'(\cdot)$ is an anisotropic, scaled, and rotated Gaussian function. We convolve the LoG filtered image with filters of varying wavelength (λ), scale (Σ), and orientation (θ), and keep the maximum response at each pixel:

$$G(\mathbf{p}) = \max_{\lambda,\Sigma,\theta}(L(\mathbf{p}) * B(\mathbf{p};\lambda,\Sigma,\theta)). \tag{16}$$

5. Image mosaicing

Image mosaicing refers to the process of fusing two or more partially overlapping images into a composite whole with a larger FOV [32]. Mosaicing consists of three distinct steps: (1) spatial registration, (2) color mapping, and (3) pixel selection or blending. We address each subproblem in the following subsections in turn.

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

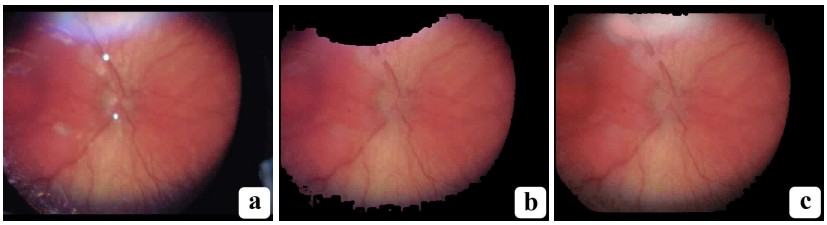

Fig. 7. Distorted color adjustment: (a) Frame affected by distorted colors arising from the lens's optics. (b) HSV masking without distorted color adjustment (c) HSV masking with distorted color adjustment. Note that the retinal area is significantly larger in (c) than in (b).

5.1. Frame registration

Standard registration techniques [1, 33, 34, 35, 36, 37] produce significant alignment errors when applied to VIO frames directly: dense methods fare poorly due to low contrast and large inter-frame motion, and feature-based methods can be biased by more visually salient but anatomically irrelevant parts of the raw images, such as the lens' rim and the various speckles.

To address these problems, we developed a three stage registration method that takes advantage of the high contrast and lack of artifacts in the Gabor vessel maps. As discussed in detail in the following paragraphs, in the first stage a global L_2-norm fit computes an approximate rigid transformation between source and target Gabor images. The second stage refines this initial estimate by matching pixels locally in high-contrast, anatomically relevant regions of the Gabor image using local rigid transformations. Finally, from the set of matched point pairs, we estimate a full affine transformation using robust L_1-norm minimization. Unlike ICP-based approaches [38, 39], our method requires only a single execution of the three stages, without any iteration. The three stages are described next.

5.1.1. Global fit

In this stage, two Gabor filtered images G_t and G_s are aligned through a rigid transformation

$$\mathbf{d}_b, \theta_b = \underset{\mathbf{d},\theta}{\operatorname{argmin}} \underset{\mathbf{d},\theta}{\mathrm{SSD}}, \tag{17}$$

where the sum-of-squared differences (SSD) for a translation $\mathbf{d}$ and rotation θ is given by:

$$\underset{\mathbf{d},\theta}{\mathrm{SSD}} = \sum_{\mathbf{p}\in P} (G_t(\mathbf{p}) - G_s(R\mathbf{p}+\mathbf{d}))^2, \quad R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{18}$$

We efficiently minimize the SSD in the frequency domain [40, 41]. Specifically, we expand Eq. (18) into three components:

$$\underset{\mathbf{d},\theta}{\mathrm{SSD}} = \sum_{\mathbf{p}\in P} (G_t(\mathbf{p}))^2 + \sum_{\mathbf{p}\in P} (G_s(R\mathbf{p}+\mathbf{d}))^2 - 2\sum_{\mathbf{p}\in P} G_t(\mathbf{p}) G_s(R\mathbf{p}+\mathbf{d}). \tag{19}$$

The first term does not depend on either $\mathbf{d}$ or θ, so we ignore it. The last term is a cross-correlation, and the second is the sum of squares of the pixels in the warped G_s that overlap G_t. For a fixed θ, we determine the last two terms for every possible translation using phase correlation:

$$\underset{\theta}{\mathrm{PSSD}} = \mathscr{F}^{-1}\{\mathscr{F}\{(G_s(R\mathbf{p}))^2\}\mathscr{F}\{U(\mathbf{p})\}\} - 2\mathscr{F}^{-1}\{\mathscr{F}\{G_t(\mathbf{p})\}\mathscr{F}\{G_s(R\mathbf{p})\}\}, \tag{20}$$

where $\mathscr{F}$ is the Fourier transform and U is an uniform weighing function over G_t. Due to frame heterogeneity, using polar or log-polar coordinates to estimate the rotation between frames [42, 43] does not yield good results in VIO. Instead, we minimize the above equation for a set of possible rotations $\Theta = \{\theta_1, \theta_2, ..., \theta_Q\}$:

$$\theta_b = \underset{\theta_i}{\operatorname{argmin}} \underset{\theta_i}{\mathrm{PSSD}}, \quad i \in [1, Q], \tag{21}$$

from which we obtain $\mathbf{d}_b = \operatorname{argmin}_{\mathbf{d}} \mathrm{PSSD}_{\theta_b}$.

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

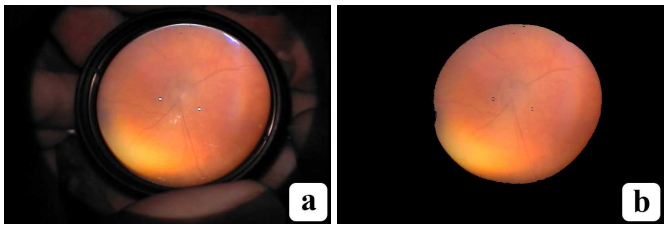

Fig. 8. HSV masking: Pixels in (a) that fall outside the HSV boundary S are flagged and discarded from further processing. Non-retinal pixels are shown as black in (b).

5.1.2. Local fit

The optimal SSD rigid-body warping removes large differences between the two frames. However, other deformations between the two vascular networks remain due to the condensing lens's optics and eye motion. We use a feature-based approach to refine the rigid-body warping into a full affine transformation. This refinement consists of two steps: correspondence point matching and transform estimation. Inferring a rigid transformation requires at least two matched pixel pairs [1]; we determine between 20 and 30 pairs for both accuracy and robustness. Our initial rigid-body estimate provides a good starting value that reduces the likelihood of converging to a local optimum.

To estimate the set of matching pixel pairs, we use the Gabor responses to determine a set of K small pixel windows $\mathbf{W}^s = \{W_1^s, W_2^s, ..., W_K^s\}$ in the source image. We then find a matching window W_k^t in the target image for each W_k^s in the source image. These windows act as corresponding features between the two images [44].

Gabor-filtered images are highly simplified: non-vessel pixels are smoothed to a uniform value and vessels are highly concentrated near an intensity extremum. This lack of texture reduces the effectiveness of dense pixel methods and general-purpose feature detectors, such as SIFT [20]. However, Gabor-filtered images encode the maximum filter response at each pixel. The pixels with the strongest responses consistently belong to the most prominent vessels in the image. Because of this, we proceed as follows to find good locations in the source image:

The center pixel $\mathbf{c}_1^s$ of the first window W_1^s is the pixel with the strongest Gabor response:

$$\mathbf{c}_1^s = \underset{\mathbf{p}\in P}{\operatorname{argmax}}\, G_s(\mathbf{p}). \tag{22}$$

For subsequent window centers, and to ensure that windows are not clustered together, we mask out from further consideration pixels that lie too close to previously selected centers:

$$f(\mathbf{p}) < t_\mathrm{d} \Rightarrow G_s(\mathbf{p}) \leftarrow \phi\ . \tag{23}$$

Here,

$$f(\mathbf{p}) = \min_{\mathbf{c}\in C_s} ||\mathbf{p}-\mathbf{c}||_2 \tag{24}$$

is the shortest Euclidean distance from a previously chosen window center, $C_s = \{\mathbf{c}_1^s, \ldots, \mathbf{c}_{k-1}^s\}$ is the current set of window center pixels, t_d is a distance threshold, and ϕ is a value that indicates that a pixel is invalid. We then determine subsequent center pixels from the part of image that has not been masked away:

$$\mathbf{c}_k^s = \underset{\mathbf{p}\notin P^\phi}{\operatorname{argmax}}\, G_s(\mathbf{p}), \tag{25}$$

where P^ϕ is the set of masked pixels. As Fig. 9 (c) shows, this procedure distributes the points throughout the image, and places them mainly along the most prominent vessels.

For each source window W_k^s thus found, we compute the best corresponding window W_k^t in the target image by the FFT SSD method outlined in Subsection 5.1.1, thereby determining the most likely rotation and translation for each window pair:

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

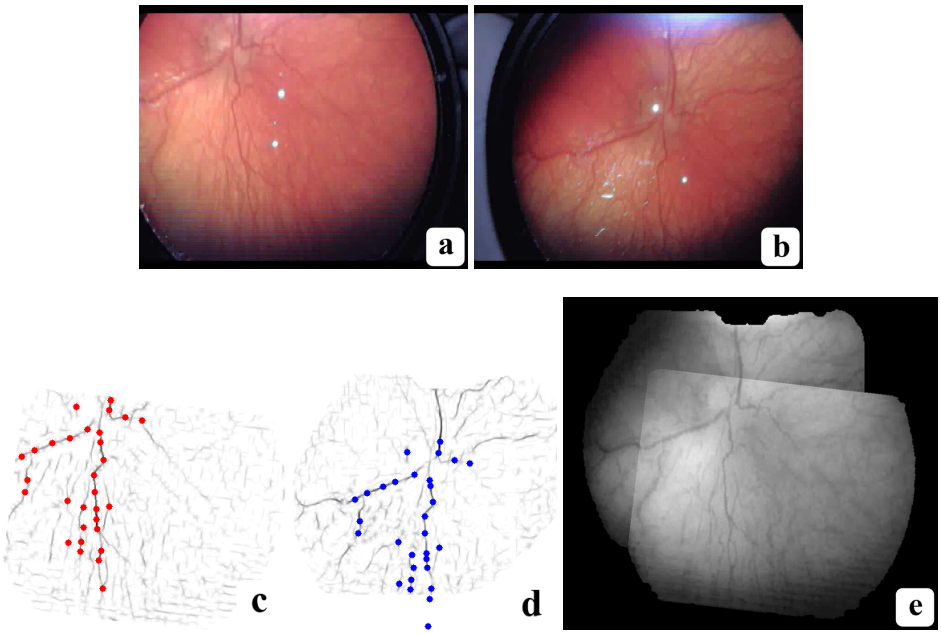

Fig. 9. Registration between two frames: (a) Source frame. (b) Target frame. (c) Source correspondence points on Gabor image. (d) Matched points on target image. Poorly matched source points are discarded. (e) Registered green channel overlay. The source frame is correctly aligned with the target frame.

$$\mathbf{d}_k, \theta_k = \underset{\mathbf{d},\theta}{\operatorname{argmin}} \sum_{\mathbf{p}\in W_k^s} (W_k^s(\mathbf{p}) - W_k^t(R\mathbf{p}+\mathbf{d}))^2. \tag{26}$$

The matching center pixel in the target image is then

$$\mathbf{c}_k^t = R_k \mathbf{c}_k^s + \mathbf{d}_k. \tag{27}$$

5.1.3. Global affine match

The third stage estimates the global affine transformation between the two images by packaging the K point matches into two $3\times K$ matrices $\mathbf{C}_s'$ and $\mathbf{C}_t'$ containing the homogeneous coordinates $\mathbf{c}' = [\mathbf{c}, 1]^T$ of the points to be aligned, and solving the following optimization problem:

$$A_{L_1} = \underset{A}{\operatorname{argmin}} \|A\mathbf{C}_s' - \mathbf{C}_t'\|_1. \tag{28}$$

Image noise, poor contrast and the occasional absence of corresponding retinal structures in the target image can cause poor feature matching. These outliers preclude the use of ordinary least squares minimization for Eq. (28) [45, 46]. Instead, we employ a robust L_1-norm minimization to obtain A_{L_1}. We iteratively converge on A_{L_1} by using the biconjugate gradient stabilized method [47, 48] on the absolute difference $\|A\mathbf{C}_s' - \mathbf{C}_t'\|_1$ to solve for the unknown affine parameters in the 3×3 matrix A_{L_1} [49].

5.2. Color mapping

VIO frames exhibit global variations in illumination that need to be compensated for prior to the final mosaicing. As illustrated in Fig. 5 (a), without color adjustment, inter-frame boundaries can form artificial border artifacts. To address this problem, we align the color space of each frame I to that of the current mosaic J by solving an absolute orientation problem in color space. Absolute orientation seeks the translation $\mathbf{d}_a$ and rotation matrix R_a that minimize the SSD between two matched point sets in $\mathbb{R}^3$ space under rigid body transformations [50]:

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

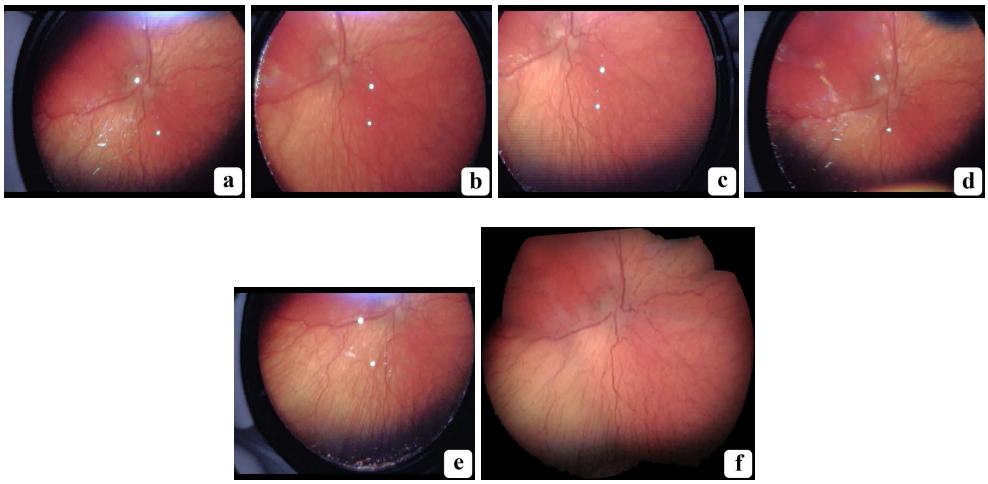

Fig. 10. VIO mosaic: A mosaic generated from five frames selected from a single VIO video. (a)-(e): each original frame; (f): the five-frame mosaic. Note the larger FOV of the mosaic relative to each individual frame, and the lack of artifacts in the final image. All images are at the same scale, and are best viewed on screen.

$$\mathbf{d}_a, R_a = \underset{\mathbf{d},R}{\operatorname{argmin}} \sum_{\mathbf{p}\in P} ||J(\mathbf{p}) - (RI(\mathbf{p}) + \mathbf{d})||_2^2. \tag{29}$$

Equation (29) has a similar form to Eq. (18), but with two key differences: First, R_a and $\mathbf{d}_a$ are applied to the image values (three-dimensional vectors), not the positions. Second, the operation is on the RGB color images I and J, not the grayscale Gabor filtered image G. We use Horn's closed form solution [51] to align the color spaces between two frames.

5.3. Pixel selection

Spatial registration and color mapping minimize geometric and photometric inter-frame differences, placing all pixel positions and values in a common frame of reference. We select the actual mosaic pixel values by a two step process consisting of feathering the pixels close to each frame's retinal boundary followed by Gabor-weighted color maximization, as described next.

5.3.1. Retinal boundary feathering

A successful mosaic minimizes inter-image seams. To reduce the visual impact of image borders, we attenuate pixel values near the edge of the valid retinal data by weights related to their Euclidean distance to the closest non-retinal pixel:

$$w_{\mathrm{E}}(\mathbf{p}) = \min_{\mathbf{p}_i \notin P_R} ||\mathbf{p} - \mathbf{p}_i||_2, \tag{30}$$

where P_R is the set of retinal pixels in I. This weighting is also known as feathering [32]. However, using the Euclidean distances as weights directly produces excessive gradation. We apply a logistic sigmoid function to w_{E} to enforce feathering only near the retinal boundary:

$$w(\mathbf{p}) = \frac{1}{1 + e^{-\lambda w_{\mathrm{E}}(\mathbf{p})}}, \quad \lambda > 0 \tag{31}$$

where λ controls the speed of weight decay. Since $w_{\mathrm{E}}(\mathbf{p})$ is non-negative, $w(\mathbf{p}) \in [0.5, 1]$.

5.3.2. Gabor-weighted color maximization

For the final pixel value selection, we weigh the color data from each frame by the corresponding Gabor responses. We determine the final mosaic in two steps. First, we compute the

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

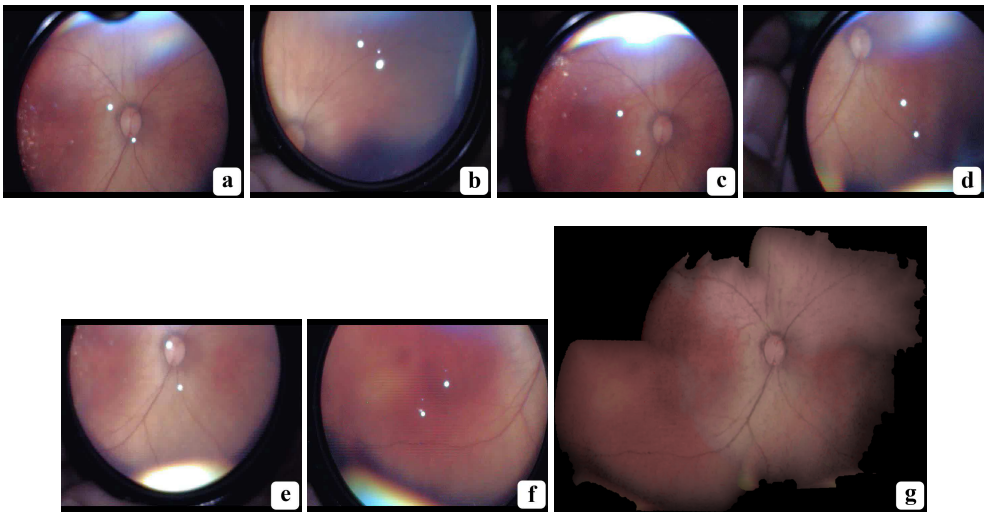

Fig. 11. VIO mosaic: A mosaic generated from six frames selected from a single VIO video. (a)-(f): each original frame; (g): the six-frame mosaic. Note the larger FOV of the mosaic relative to each individual frame, and the lack of artifacts in the final image. All images are at the same scale, and are best viewed on screen.

brightest value of each color vector's components independently at each pixel position, and across all N' images to be merged:

$$M_C^h(\mathbf{p}) = \max_n I_n^h(\mathbf{p}), \quad n \in [1, N'], \tag{32}$$

Here, I has been spatially and photometrically registered, and the h superscript indicates an RGB color channel. We then construct a complement Gabor mosaic separately:

$$\hat{M}_G(\mathbf{p}) = \min_n \hat{G}_n(\mathbf{p}), \quad n \in [1, N']. \tag{33}$$

As noted in Section 4, in a complement image, the image values are inversely proportional to the filter responses. We use the complement Gabor images in order to keep vessels darker than the surrounding tissue. The final mosaic is a convex combination of the two mosaics:

$$M = \alpha M_C + (1-\alpha)\hat{M}_G. \tag{34}$$

In Section 6, we empirically determined the best value of α for the quantitative experiments. Figures 10 and 11 show two mosaics constructed using this method. They illustrate that our proposed pipeline can convert raw videos of the quality represented by Fig. 4 into seamless mosaics of a neonate's retina. These mosaics not only extend the FOV of the original frames, but significantly reduce the artifacts and spurious data found in the constituent images.

6. Clinical experiments

The mosaics of Fig. 10 and 11 verify that our pipeline produces wide FOV mosaics with minimal artifacts that are suitable for human visual inspection. Nevertheless, while manual IO is the standard of care for ROP, quantitative approaches to ROP and plus disease diagnosis are becoming increasingly important. It is therefore crucial for our methodology to be suitable for automated and semi-automated processing as well. As a motivating example, we first examined both of these mosaics and the best, hand-picked frame for each source video with ROPTool. As Fig. 12 shows, the mosaics allowed us extract longer, more numerous vessels in less time compared to the single frames.

To more broadly verify the clinical usefulness of our proposed approach, we carried out a pilot study using ROPTool [6] in which we compared our automatic mosaicing pipeline to the

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

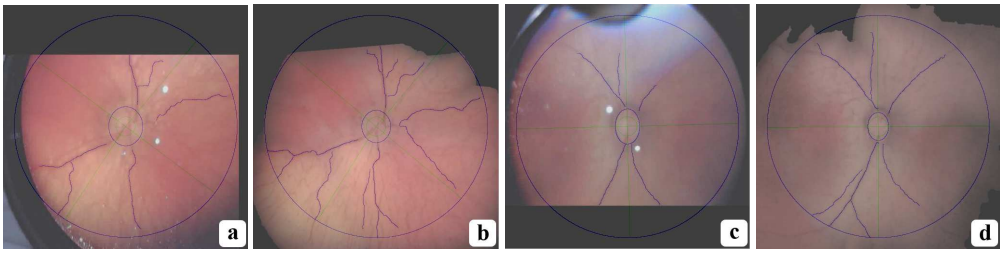

Fig. 12. ROPTool analysis comparison: ROPTool analysis of the mosaics of Fig. 10 and 11 as well as the best hand-picked frames from the corresponding videos: (a), (c) Hand-picked frames; (b), (d) Mosaics. Images are best viewed on screen. The blue lines inside the large blue circle indicate the vessel paths obtained with ROPTool. At least one, but preferably two major vessels from each quadrant are needed to provide a full ROP diagnosis. ROPTool is able to extract longer and more numerous vessels when analyzing the mosaics than the hand-picked frames. Furthermore, ROPTool analysis of the mosaics is faster due to fewer ROPTool mistakes. The examination times (in minutes) were: (a) 2:45, (b) **2:00**, (c) 2:35, (d) **1:30**

current state-of-the-art procedure for semi-automatic ROP analysis. This procedure relies on manually selected frames captured from a live feed during the actual IO examination. First, while the physician carries out the IO examination, the recorded video feed is displayed on a bedside computer. Meanwhile, an assistant observes the live feed and manually selects a few dozen potentially good frames by clicking on a frame capture button. The physician later sifts through the captured frames and selects the two best frames, one for each eye. Finally, the physician performs the ROPTool analysis on the two selected frames.

There are a number of drawback to this procedure. The first is that it requires careful manual examination of the recorded video feed. Any loss of concentration by the assistant can result in valuable frames been overlooked. Second, the frame capture operation is approximate due to two sources of lag: (1) the time between when the assistant sees a good frame and when he or she clicks the screen capture button, and (2) the time between which the computer registers the mouse click and when it performs the frame grabbing operation. Furthermore, even if the assistant manages to select the single best frame in a video, it can still be affected by limited FOV and numerous imaging artifacts. Finally, even in this best-case scenario, most of the video's information is discarded. In the following three subsections, we first detail the methods used to obtain and analyze the VIO data, then describe the exact mosaicing steps and parameters used, and finally present and discuss the ROPTool results.

6.1. Methods

We obtained 31 VIO videos from six ROP examination sessions performed by two expert pediatric ophthalmologists on 15 different patients at the Neonatal Intensive Care Unit at the Duke University Medical Center in Durham, NC. All the videos were recorded using a Keeler Wireless Digital Indirect Ophthalmoscope (Keeler, Instruments Inc, Broomall, PA). The videos in all the sessions were captured at a resolution of 720×576 pixels in 24-bit color and saved as Audio Video Interleaved (AVI) files. During each IO examination, an assistant observed a live version of the video feed at 30 fps and manually selected several dozen frames per video. After the IO examination, we then separately constructed a full mosaic from each video with our proposed method. For consistency, all manually selected images and mosaics are of each patient's right eye.

Dr. Cabrera, an expert pediatric ophthalmologist, independently carried out a ROPTool analysis of both types of images. She first selected the best manual frame for each right eye from the several dozen captured by the assistant and then performed the ROPTool analysis on this frame and the corresponding mosaic. She first located the optic nerve. Based on its location, ROPTool divided the image into four quadrants. She then clicked on a point within each visible vessel in

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

Table 1. Quantitative ROPTool analysis of 31 mosaics and their corresponding best hand-picked frames

Image type	Traceable	Diagnosis agreement	Tracing time (in minutes)
Mosaics	30 (96.8%)[a]	24 (80.0%)[b] (77.4%)[a]	1:41 ($\pm$ 0.03)
Handpicked images	26 (83.9%)[a]	19 (73.1%)[b] (61.3%)[a]	2:42 ($\pm$ 0.04)

[a] Percentage of total frames (31).
[b] Percentage of traceable frames.

the image and ROPTool traced the rest of the selected vessel. She then determined if the image was traceable, based on the criteria detailed in [6]. In summary, a vessel is deemed traceable if it can be traced for a distance of at least one optic disc diameter outward from its junction with the edge of the optic nerve and a quadrant is traceable if there is at least one traceable vessel in it. An image as a whole is traceable if at least three out of four quadrants are traceable. ROPTool then provided a diagnosis based on the tortuosity and dilation of the traced vessels. She then recorded the examination time and compared ROPTool's diagnosis based on the image with the diagnosis given by either Dr. Wallace or Dr. Freedman during the IO examination.

6.2. Implementation details

To obtain a mosaic from a VIO video, we first empirically determined suitable parameter values for the dataset based on the mosaics in Fig. 10 and 11. We used the same set of parameters for all videos. For every video, we then applied HSV classification to every frame using an interactively chosen retinal pixel value $\mathbf{v}$. We converted this RGB pixel value into HSV: $\mathbf{v}_{\mathrm{HSV}}$ and set it as the center of the closed surface S. We only needed one pixel for every video. We constructed S by forming three intervals: $[\mathbf{v}^h_{\mathrm{HSV}} - t_h, \mathbf{v}^h_{\mathrm{HSV}} + t_h], [\mathbf{v}^s_{\mathrm{HSV}} - t_s, \mathbf{v}^s_{\mathrm{HSV}} + t_s], [\mathbf{v}^v_{\mathrm{HSV}} - t_v, \mathbf{v}^v_{\mathrm{HSV}} + t_v]$. The value for t_h, t_s and t_v was 0.1.

The frames with the highest 10% of HSV scores (generally 100 to 300 frames per video) were then analyzed using the spatial frequency measure. We used smoothing parameters $\sigma_m = 0.25$ and $\sigma_l = 0.1$. We obtained a convex combination score of the two measures using $\upsilon = 0.3$, in Eq. (7). The top 20% to 50% of those frames–2% to 5% of total frames, about 20 frames per video–were saved as TIFF files. Retaining a higher percentage of frames was not needed since frames with lower percentile scores were rarely of interest. From these 20 images, a small number (usually 4 to 6) were manually selected per video, in order to ensure that they all corresponded to the right eye. For each of the selected frames, we applied directional local contrast filtering with a removal threshold of 1.1 above the frame's median value. We applied multiscale LoG ($\sigma = [0.11, 0.51]$, with step-size $\sigma_\Delta = 0.1$) and Gabor wavelet filtering ($a = 10, b = 10, u = 0.05, v = 0.05, \beta = 30, \theta = [0, 170]$, with step-size $\theta_\Delta = 10$) for vessel mapping. The most central frame was selected as the target frame and the remaining frames were registered in two stages. We determined a global fit by iteratively searching over a range of r rotations with r_Δ stepsize. We used values $r = [-\pi/12, \pi/12]$ and $r_\Delta = 3$. For each rotation, we determined the optimal translation using phase correlation. We selected 30 control points on the source image, each with a 20 pixel buffer zone, and locally registered 10×10-pixel windows around each one. We estimated the full affine matrix by applying gradient descent on the sum of absolute differences between the matched pixel pairs. We used an update weight γ of 0.01 and a maximum of 5000 iterations. We used absolute orientation to map every source frame's color to the same target frame. We used a λ of 0.1 for feathering the borders of the valid retinal regions in each frame. Finally, the mosaic was composed using Gabor-weighted color maximization, treating each channel separately. To determine the optimal value for α, Dr. Cabrera analyzed a number of mosaics using ROPTool for three settings of α: $[0.25, 0.5, 0.75]$. Her analysis established that $\alpha = 0.75$ produced the best results.

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011

6.3. Results

Table 1 presents the results from our pilot ROPTool study, in which we analyzed 31 hand-picked frames and 31 corresponding mosaics from the right eyes of 15 neonate patients. As Dr. Cabrera observed, the automatically constructed mosaics allowed ROPTool to extract longer vessels with fewer tracing errors compared to even the best hand-picked frames. Most importantly, ROPTool's diagnosis was more often in agreement with the ROP experts' diagnosis for our mosaics than for the hand-picked frames. The difference is particularly striking for the percentage of correctly diagnosed images relative to the total number of images (77.4% vs. 61.3%); the total number of images was 31 in both cases. This value represents how often ROPTool is clinically valuable and suggests that our mosaicing pipeline improves the diagnostic accuracy of semi-automated ROP analysis by almost 20%. Finally, in spite of being able to trace more vessels, our mosaics allow the operator to reduce the tracing time by an average of 37%.

7. Conclusions

We have developed an effective and efficient pipeline for constructing a high quality, large FOV mosaic from a raw VIO video. The mosaic is suitable for human or machine analysis and diagnosis. Our initial results suggest that the use of mosaics allows semi-automated ROP analysis programs such as ROPTool to better match the diagnostic assessments of human experts. By removing artifacts, mapping vessels in different frames and fusing the automatically selected best frames, we were able to overcome the numerous complications that arise from VIO recording.

The next step in our research will involve further validating the diagnostic utility of these mosaics by means of a larger scale clinical study. We will more precisely determine the statistical significance of using our mosaics for accurate human and semi-automated ROP diagnoses.

The algorithms presented in this paper were mainly discussed for the specific task of VIO retinal imaging. However, the mathematical techniques are general and can be modified to enhance the quality of other en-face imaging modalities. Moreover, the 2-D methodology presented here may be generalized for creating 3-D mosaic. Generalizing our methods to confocal scanning laser ophthalmoscopy and spectral domain optical coherence tomography is part of our ongoing work [52].

Acknowledgments

This work was supported in part by the Knights Templar Eye Foundation, Inc. and the Hartwell Foundation.

#152575 - $15.00 USD Received 8 Aug 2011; revised 16 Sep 2011; accepted 17 Sep 2011; published 29 Sep 2011