

Autonomous Stabilization of Retinal Videos for Streamlining Assessment of Spontaneous Venous Pulsations

Hongwei Sheng^{1,2}, Xin Yu², Feiyu Wang¹, MD Wahiduzzaman Khan^{1,2}, Hexuan Weng¹, Sahar Shariflou¹, S.Mojtaba Golzan¹

Abstract—Spontaneous retinal Venous Pulsations (SVP) are rhythmic changes in the caliber of the central retinal vein and are observed in the optic disc region (ODR) of the retina. Its absence is a critical indicator of various ocular or neurological abnormalities. Recent advances in imaging technology have enabled the development of portable smartphone-based devices for observing the retina and assessment of SVPs. However, the quality of smartphone-based retinal videos is often poor due to noise and image jitting, which in return, can severely obstruct the observation of SVPs. In this work, we developed a fully automated retinal video stabilization method that enables the examination of SVPs captured by various mobile devices. Specifically, we first propose an ODR Spatio-Temporal Localization (ODR-STL) module to localize visible ODR and remove noisy and jittering frames. Then, we introduce a Noise-Aware Template Matching (NATM) module to stabilize high-quality video segments at a fixed position in the field of view. After the processing, the SVPs can be easily observed in the stabilized videos, significantly facilitating user observations. Furthermore, our method is cost-effective and has been tested in both subjective and objective evaluations. Both of the evaluations support its effectiveness in facilitating the observation of SVPs. This can improve the timely diagnosis and treatment of associated diseases, making it a valuable tool for eye health professionals.

I. INTRODUCTION

Spontaneous retinal venous pulsations (SVP) are rhythmic changes in the central retinal vein and its branches. SVPs are commonly present in the optic disc region (ODR) of the retina. Absent SVPs are clinically associated with progression in glaucoma and increased intracranial pressure [1], [2]. Accordingly, assessment of the retina in determining SVP presence is clinically paramount.

SVP evaluation is performed by inspecting the deformation of the retina vessels via fundus photography techniques [3]. Conventional fundus photography data is usually captured using specialized and expensive benchtop equipment operated by trained professionals. However, with increasing emphasis on eye health, this method is not sufficient to meet the growing demand.

Recently, due to the low cost and easy accessibility of smartphones, researchers and clinicians [4], [5] have commenced using smartphone-based fundus photography to

*This research is funded by ARC-Discovery grant(DP220100800) and ARC-DECRA grant(DE230100477).

¹Hongwei Sheng, MD Wahiduzzaman Khan, Hexuan Weng, Feiyu Wang, Sahar Shariflou and S.Mojtaba Golzan are with the University of Technology Sydney, NSW 2007, Australia

²Hongwei Sheng, MD Wahiduzzaman Khan, Xin Yu are with the University of Queensland, QLD 4067, Australia

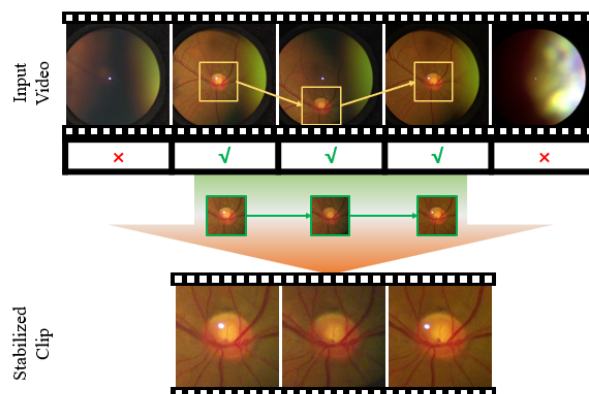


Fig. 1. Illustration of our proposed automatic fundus retina video stabilization method. Given a fundus retina video, our method automatically detects the ODR both spatially and temporally. Afterward, our approach aligns the visible ODR clips to a fixed position in the video footage. The stabilized video facilitates better SVP inspection.

assess retinal conditions, allowing frequent SVP observation. However, retinal fundus images captured by hand-held devices are not robust to various real-world artifacts, such as noise and jittering. These artifacts would make medical analysis difficult and time-consuming [6], [7], [8], thus presenting challenges in clinical diagnosis and exacerbating the likelihood of erroneous evaluation.

Most existing video stabilization works are proposed to stabilize natural scenes [9] without taking specific image domain knowledge into account. Only a few works [10], [11] have been proposed to stabilize fundus retina videos. However, those methods often require high-quality videos. For example, the videos cannot contain eye blinks and drastic illumination changes. Thus, those methods are not suitable to process mobile-captured real-world retina videos. Efficiently monitoring SVPs with computer-assisted techniques [12], [13] still remains challenging. It is necessary to design an effective automatic mobile-based fundus video stabilization method to enable easy observation and diagnosis.

To better support clinicians or even non-experts in observing and evaluating SVPs, we propose a fully automatic fundus retina video stabilization method as shown in Fig. 1.

Specifically, we design an ODR Spatio-Temporal Localization (ODR-STL) module to first detect the spatial locations of ODR from a fundus video and then temporally localize the ODR-visible clips from the video.

Next, we introduce a Noise-Aware Template Matching (NATM) module to stabilize the ODR-visible video clips to a fixed position of the footage (*i.e.*, the field of view). In the

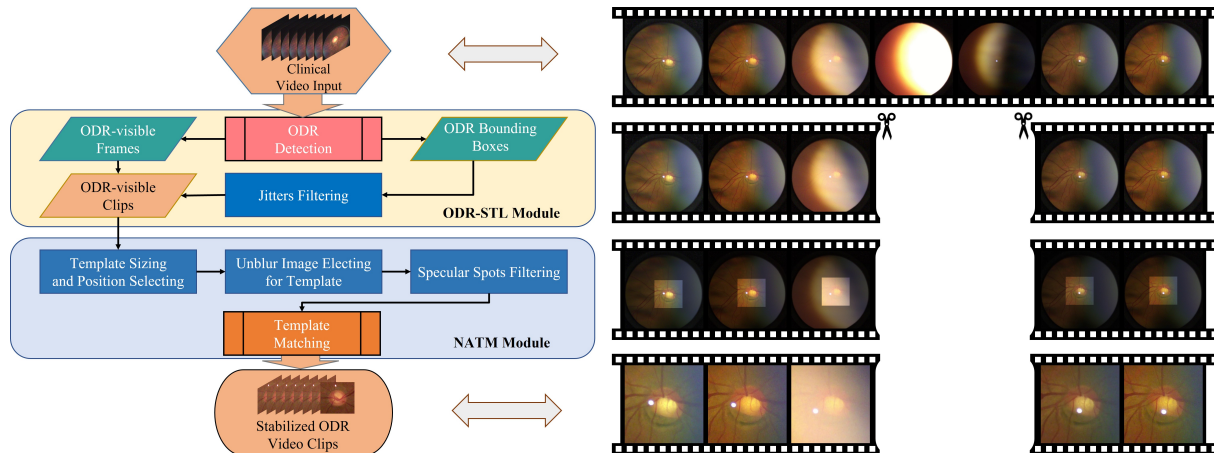


Fig. 2. Overview of the pipeline of our method. Left: The video will be first processed by an ODR Spatio-Temporal Localization Module to remove ODR-invisible frames. Then, a Noise Aware Template Matching is applied to align high-quality ODR frames to a fixed position. The video is cropped to focus on ODR regions, thus aiding SVP inspection. Right: The intermediate results of our processing pipeline.

meanwhile, NATM is robust to specular near ODR regions since specular might severely affect conventional template matching performance.

After the stabilization, the observation of ODR videos will not be subject to the artifacts caused by eyeball movements, blinks, or abrupt brightness fluctuation, thus significantly facilitating clinical assessment and downstream machine perception. To further illustrate the effectiveness of our proposed method, we apply it to real fundus videos captured with various types of smartphone-based fundus cameras. Moreover, we quantitatively evaluate the quality of the stabilized video clips as well as conduct a user study in which clinicians are invited to choose which videos are more favored in observing and evaluating SVPs.

II. METHODOLOGY

Our fully automated stabilization method is designed to facilitate the detection of SVPs from clinical fundus videos. As shown in Fig. 2, we feed an input fundus retina video captured by the smartphone-based retina imaging platform in clinics to our system and then obtain a stabilized video of optical disc regions, where SVPs usually emerge. In our method, the first step is a spatio-temporal localization of visible ODR (ODR-STL) and then we select continuous ODR-visible frames. The second step is to apply noise-aware template matching (NATM) to stabilize the positions of visible ODR in video clips.

A. Spatio-temporal Localization of Visible ODR

To accurately determine the presence of SVP, clinicians need to observe at least one full clear cycle of the pulsation in a fundus video. Thus, our ODR-STL method first detects frames that contain ODR in the footage. Then our algorithm identifies and eliminates huge jitters in the footage. As a result, video clips that contain continuous ODR with sufficient temporal duration (*i.e.*, longer than a full cycle of SVP) are kept for further stabilization.

1) *ODR Detection*: Our method first employs a faster R-CNN neural network [14] to detect ODR in video frames. Utilizing deep networks helps to improve the accuracy of the detecting results and circumvents some limitations of traditional methods (*e.g.*, sensitive to noise and illumination changes). With the faster R-CNN model, our method extracts the bounding box of ODR in each frame as well as obtains the existence of ODR along the temporal dimension. As a result, we obtain the spatial and temporal information of visual ODR in a video.

2) *Filtering Jitters*: Our method then utilizes bounding boxes to eliminate the jitters in the footage. Due to the inherent noise presented in the video data and the size variance of detected boxes, the detected bounding boxes present frequent fluctuations. However, since the fluctuation around the ODR is small compared to the sizes of bounding boxes, our method still obtains the approximate location of the ODR in each frame, enabling further screening extreme jitters. As shown in Fig. 3, the polylines reflect the movements of ODR in the footage throughout the video. The broken parts of the polyline indicate that the ODR is not detected in these frames. By eliminating the huge jitter marked as yellow vertical, our method further preserves a continuous clip with ODR consistently showing in footage.

B. Noise-Aware Template Matching

Our NATM method is then applied to register the positions of ODR in each clip. Due to the presence of random noise, it is unsuitable to apply the plain template matching algorithm to the data. Therefore, three additional strategies have been introduced to mitigate the impact of noise when applying template matching.

1) *Template Size Selection*: Template-based matching requires a template as a reference. The algorithm identifies the region of interest by evaluating the pixel-wise difference in RGB values between a template and a potential matching target. Thus, it is important to choose a template from a region that is significantly different from the other parts of the image. To achieve the best matching results on clinical noisy data, the size and content of the template are very

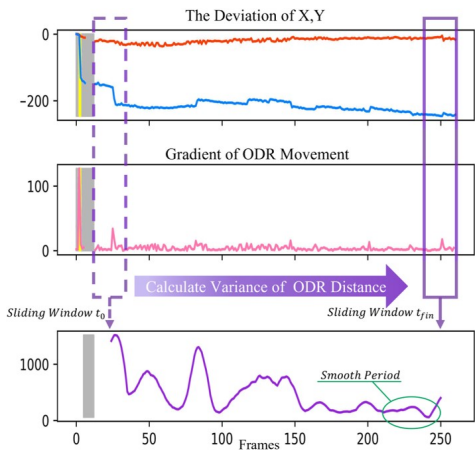


Fig. 3. ODR Trajectories. The red (X-axis) and blue (Y-axis) lines indicate detected ODR deviated from the position in the first frame. The gray regions indicate the removed frames after localization. The pink line indicates the gradient of ODR trajectories. The purple line indicates the variance of ODR distance over frames.

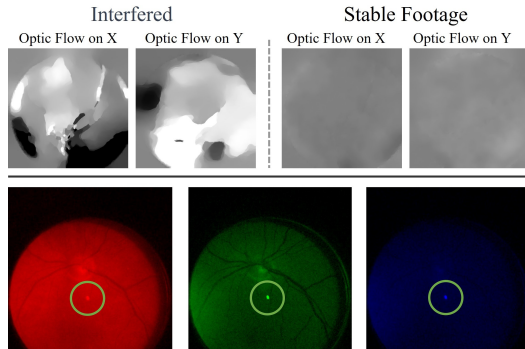


Fig. 5. Top: Optical flow before and after stabilization. Top left: the optical flow of an original video along X and Y dimensions, respectively. Top right: the optical flow of our stabilized video. The smoother flow indicates that the video has fewer jittering artifacts. Bottom: Impacts of specular on different channels in template matching. We can observe that the blue and green channels are very sensitive to specular. Therefore, in order to achieve a noise-aware template matching, our method reduces the impacts of blue and green channels on template matching.

crucial. Since ODR is the brightest area and is unique in retina images, the best choice for our task is to employ a rectangle template centered on the ODR, with the side length overlapping the ODR diameter. The ODR diameter is estimated from the sizes of bounding boxes detected by the faster R-CNN.

2) *Screening Blur Template*: After confirming the size and position of the template, our method then determines which frame should be adopted as the template. In order to accurately find the ODR, our method selects the template from a video frame without blur. Considering the irregular noise in a video, it is difficult to calculate sharpness via conventional gradient-based algorithms such as Laplacian sharpness measurement. Meanwhile, we notice that blur frames usually enlarge the jitters of bounding boxes. Therefore, in each video clip, we use the trajectory of bounding box centers within a sliding window to select the most smooth period of the video, as shown in Fig. 3. Then, we

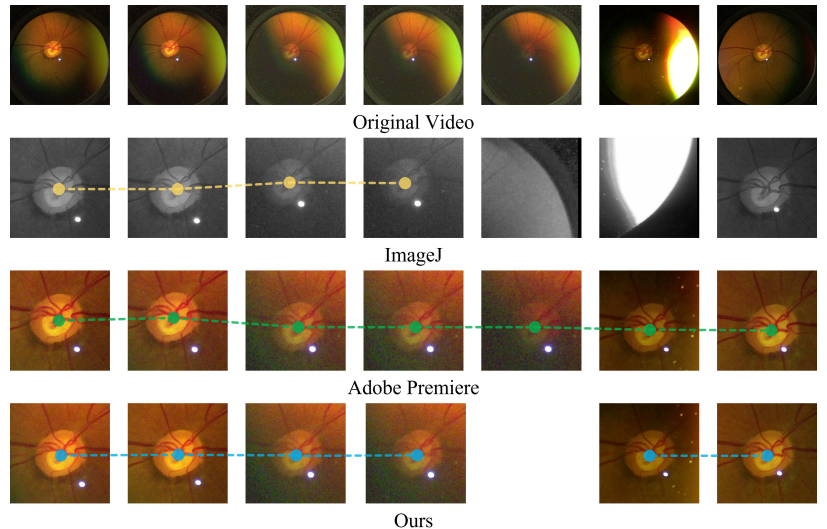


Fig. 4. Comparison with other video stabilization methods. Note that our method is able to remove low-quality images. In this example, ODR does not have enough illumination and thus we remove this frame and obtain two video clips.

use the variance of optical flow [15] to determine the quality of each frame in terms of sharpness as shown in Fig. 5. The frame with the lowest variance of the optical flow will be considered as the sharpest image in the most smooth period of the clip.

3) *Specular Spots Removing*: In practice, some hand-held fundus video-capturing devices utilize external light sources to better observe fundus retina. This results in specular spots on the eyeballs due to the reflection. As the light sources normally emit white light, the reflected specular spots present high values in RGB channels as shown in Fig. 5. We can observe that specular spots protrude especially in the blue (B) and green (G) channels. Thus, we selected an appropriate global threshold on the B and G channels. Then, our algorithm employs mean filtering to minimize the interference of specular light spots during template matching.

With these three strategies, template matching obtains a series of ODR coordinates precisely from each noisy video clip. This allows us to easily align these coordinates and fix the positions of ODR in the footage. Our method further crops the ODR out with a modifiable size to emphasize SVP. As a result, we obtain a set of ODR-stabilized video clips from our auto-processing pipeline.

III. EXPERIMENTS

In our study, our pipeline will automatically crop video clips of 640×640 pixels from the original videos of size 1800×1800 pixels. This not only saves the storage by reducing redundant information, but also emphasizes the area where SVPs are commonly detectable. We then produce quantitative and qualitative experiments including a user study to evaluate the performance of our fundus retina video stabilization method against some popularly used approaches.

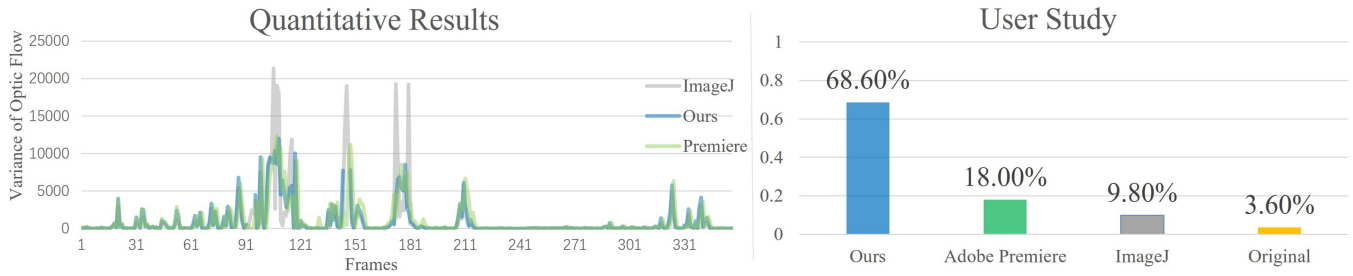


Fig. 6. Left: The variance of optic flow over frames. The variance of our method is the lowest. The variance of Adobe Premiere is close to ours, but its effort on reducing optic flow variance brings distortion to the stabilized video. Right: User Study. We invite 25 subjects including clinical professionals and non-experts to evaluate the quality of 20 stabilized videos by different methods. The original videos are also provided to the users for evaluation. Most users favor stabilized videos by our method.

A. Quantitative Results

To further illustrate the effectiveness of our method in inspecting SVPs, we conduct a series of comparative experiments as objective evaluations. We also compare with some existing video stabilization methods. ImageJ (*i.e.*, FIJI) is a widely used software in medical image processing. Its plugins provide capacity for many tasks including video stabilization. Adobe Premiere Pro is a popular video processing and editing commercial software. It has a built-in function to stabilize videos as well. We utilize optic flow to measure the stability. As shown in Fig. 6, our method achieves the least variance of optical flow.

B. Qualitative Results

Fig. 4 presents examples from our method and other methods. The original video of this example contains blur, specular spots, low illumination, and other real-world noise. We can see that ImageJ fails to address the blur and the stabilized videos suffer jittering when the specular moves. In the video clips with less noise, the performance of ImageJ can be on par with our method. Adobe Premiere Pro performs more steadily, but it introduces distortion in the stabilized video. This distortion will harm the SVP observation.

C. User Study

To determine the effectiveness of our method for manual diagnosis, we recruited 25 individual subjects with varying levels of expertise from four different clinics. We prepare some groups of test videos in advance. Each group contains four videos processed by four different methods from the same original video. We then ask the subjects to watch five groups of test videos and then report which video in each group they would like to use to best observe SVP. We graph their feedback as in Fig. 6. In this user study, videos from our method are more favored than those from the others.

IV. CONCLUSIONS

In this paper, we propose an automatic fundus retina video stabilization method that would profoundly promote the wide application of smartphone-based SVP inspection and assessment. From the stabilized videos, clinical professionals or even non-professionals can easily observe SVPs. We also believe the stabilized video can significantly ease the downstream clinical diagnosis and analysis. Noticeably,

our method has been tested on various data collected by different mobile devices in different clinics. The experiments demonstrate that our retina video stabilization method is very effective in different clinical environments.

REFERENCES

- [1] W. H. Morgan, M. L. Hazelton, and D.-Y. Yu, "Retinal venous pulsation: Expanding our understanding and use of this enigmatic phenomenon," *Progress in retinal and eye research*, vol. 55, pp. 82–107, 2016.
- [2] D. Moreno-Ajona, J. A. McHugh, and J. Hoffmann, "An update on imaging in idiopathic intracranial hypertension," *Frontiers in Neurology*, vol. 11, p. 453, 2020.
- [3] T. Hamann, M. Wiest, A. Mislavics, A. Bondarenko, and S. Zweifel, "At the pulse of time: Machine vision in retinal videos," in *Machine Learning in Clinical Neuroscience*, pp. 303–311, 2022.
- [4] M. W. Wintergerst, D. K. Mishra, L. Hartmann, P. Shah, V. K. Konana, P. Sagar, M. Berger, K. Murali, F. G. Holz, M. P. Shanmugam, *et al.*, "Diabetic retinopathy screening using smartphone-based fundus imaging in india," *Ophthalmology*, vol. 127, no. 11, pp. 1529–1538, 2020.
- [5] U. Iqbal, "Smartphone fundus photography: a narrative review," *International Journal of Retina and Vitreous*, vol. 7, no. 1, pp. 1–12, 2021.
- [6] M. Monjur, I. T. Hoque, T. Hashem, M. A. Rakib, J. E. Kim, and S. I. Ahamed, "Smartphone based fundus camera for the diagnosis of retinal diseases," *Smart Health*, vol. 19, p. 100177, 2021.
- [7] H. Liu, H. Li, X. Wang, H. Li, M. Ou, L. Hao, Y. Hu, and J. Liu, "Understanding how fundus image quality degradation affects cnn-based diagnosis," in *2022 44th EMBC*, pp. 438–442, 2022.
- [8] J. Liu and X. Yu, "Few-shot weighted style matching for glaucoma detection," in *Artificial Intelligence: First CAAI International Conference, CICA 2021, June 5–6, 2021, Proceedings, Part 1 I*, pp. 289–300, Springer, 2021.
- [9] W. Guilluy, L. Oudre, and A. Beghdadi, "Video stabilization: Overview, challenges and perspectives," *Signal Processing: Image Communication*, vol. 90, p. 116015, 2021.
- [10] P. Kingkosol, P. Pooprasert, P. Choopong, B. Hunchangsinh, V. Laksanaphuk, and C. Tantibundhit, "Automated cytomegalovirus retinitis screening in fundus images," in *2020 42nd EMBC*, pp. 1996–2002, 2020.
- [11] S. Mueller, S. Karpova, M. W. Wintergerst, K. Murali, M. P. Shanmugam, R. P. Finger, and T. Schultz, "Automated detection of diabetic retinopathy from smartphone fundus videos," in *International Workshop on Ophthalmic Medical Image Analysis*, pp. 83–92, 2020.
- [12] J. A. McHugh, L. D'Antona, A. K. Toma, and F. D. Bremner, "Spontaneous venous pulsations detected with infrared videography," *Journal of Neuro-Ophthalmology*, vol. 40, no. 2, pp. 174–177, 2020.
- [13] A. Pujari, G. Saluja, D. Agarwal, A. Sinha, A. PR, A. Kumar, and N. Sharma, "Clinical role of smartphone fundus imaging in diabetic retinopathy and other neuro-retinal diseases," *Current Eye Research*, vol. 46, no. 11, pp. 1605–1613, 2021.
- [14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [15] J. S. Pérez, E. Meinhardt-Llopis, and G. Facciolo, "Tv-11 optical flow estimation," *Image Processing On Line*, vol. 2013, pp. 137–150, 2013.